CROATIAN
JOURNAL
OF PHILOSOPHY

Vol. XX · No. 60 · 2020

On Pietroski's Conjoining Meanings

Articles

Book Discussion

Book Reviews

Introduction

The first part, namely the first five papers of this issue of the Croatian Journal of Philosophy *include the papers given at the yearly course* Philosophy of Language and Linguistics *held at the* Inter University Centre Dubrovnik *(the IUC) in September 2019. Two days of the course in 2019 were dedicated to Paul Pietroski's then forthcoming book* Conjoining Meanings, Semantics Without Truth Values*, now published by the Oxford University Press. The paper by Anna Drożdżowicz was also presented at the same conference but unrelated to Pietroski. Papers by Olga Ramirez Calle, Antonin Thuns and Rene Jagnow broadly belong to philosophy of language, too.*

The second part of this issue brings two papers from political epistemology, discussing epistemic aspect of deliberation. The paper by Nenad Miščević, discussing the work of Snježana Prijić Samaržija was planned for the previous issue, but the author was late (for which he apologizes to the readers). John B. Min's paper offers a reading of "epistemic" which brings it into constitutive relation with democratic deliberation.

DUNJA JUTRONIĆ & NENAD MIŠČEVIĆ

Croatian Journal of Philosophy
Vol. XX, No. 60, 2020

Précis of Conjoining Meanings: Semantics Without Truth Values

PAUL M. PIETROSKI*
Rutgers University, New Brunswick, NJ, USA

In Conjoining Meanings, *I argue that meanings are composable instructions for how to build concepts of a special kind. In this summary of the main line of argument, I stress that proposals about what linguistic meanings are should make room for the phenomenon of lexical polysemy. On my internalist proposal, a single lexical item can be used to access various concepts on different occasions of use. And if lexical items are often "conceptually equivocal" in this way, then some familiar arguments for externalist conceptions of linguistic meaning need to be reevaluated.*

Keywords: Meaning, polysemy, concepts, semantic internalism, Chomsky.

Children acquire languages that connect meanings with pronunciations in striking ways. I offer a proposal about what these meanings are, how they are related to human cognition, and how they are not related to the things we talk about by using words in contexts. In slogan form, meanings are composable recipes for how to build concepts of a special sort. The meaning of 'green bottle' is a tripartite instruction: access a concept via 'bottle', access a concept via 'green', and conjoin the results. Likewise, 'my green bottle' calls for conjoining the results of executing the meanings of 'my' and 'green bottle'. Full sentences can be used to build complete thoughts, which may be true or false. But in my view, ordinary sentences don't have truth values, not even relative

* This is an edited version of the précis I provided for a session on *Conjoining Meanings* (Oxford University Press, 2018) at the Philosophy of Linguistics conference in Dubrovnik in September 2019. I am enormously grateful to Dunja Jutronić for organizing the event, to the participants for helpful comments and questions, and especially to the speakers who devoted so much time and energy to preparing such thoughtful and detailed remarks in advance.

to contexts. Meanings don't determine extensions. My route to these conclusions is paved with details that surely need revision. Though at the outset, it's hard enough to identify the topic.

1. *Slangs*

Like most words, 'meaning' is polysemous. There are many kinds of significance, and many corresponding concepts. So it's easy, even for specialists, to talk past each other when using 'meaning'. But I think there is something like a natural kind in the vicinity.

The child-acquirable languages that connect meanings with pronunciations, spoken or signed, are distinctively human. Let's call them Slangs. The meanings of Slang expressions are somehow compositional in a way that allows for ambiguities yet mirrors certain aspects of logical structure. For example, speakers of English can understand (1) in several ways.

(1) we watched her duck near a muddy bank

But however 'duck' and 'bank' are construed, deleting 'muddy' or 'near a muddy bank' seems valid. Such facts indicate the meanings I have in mind: Slangs connect them with pronunciations in interesting ways. One of my central claims in *Conjoining Meanings* (*CM*) is that ambiguity and composition, along with the compelling character of certain inferences, are natural phenomena that reflect aspects of human psychology. Correlatively, we should avoid stipulations about how meanings are related to truth values, possible worlds, or extensions of ideal concepts. We have to discover the nature of the meanings that Slangs connect with pronunciations.

In principle, my claims about meaning can be combined with various suggestions about what concepts are. But chapter two offers a Fodorian account of concepts as composable mental representations with which we can think about things. This leads into a discussion of Aristotelian logic, as part of a larger argument that mental *predicates*—concepts that let us categorize—play a special role in a natural logic that vindicates an old idea: predicate reduction (e.g., replacing 'muddy bank' with 'bank') is typically *conjunct* reduction; and deleting conjuncts is valid except in specially marked environments like negation. For predicates, the default is that longer is stronger. In later chapters, I argue that phrasal meanings are instructions for how to build mental predicates, given lexical meanings that let us access a limited range of atomic concepts. But this is a hypothesis about certain natural languages, not a proposed analysis of any pretheoretic concept of meaning.

For simplicity, let's focus on spoken Slangs. And let's say that expressions with the same pronunciation are homophonous, suggesting that they "sound" the same, even if uttered by a soprano and a baritone. But let's not pretend that spoken English is a single Slang.

Pronunciations vary across speakers who grew up in Brooklyn,

Glasgow, Sydney, or different neighborhoods in London. In some cases, the same meaning is expressed with very different sounds, as with 'biscuits in the lift' vs. 'cookies in the elevator'. Conversely, the same sound—e.g., that of 'solicitor' or 'robin'—can be paired with different meanings in different places. (British solicitors refer trial work to barristers. American solicitors represent units of government and often argue cases. Similarly, British 'robin' and its American counterpart correspond to different species.) English Slangs also exhibit minor syntactic differences. And as games of Scrabble can reveal, many entries in the O.E.D. are not words of *my* Slang.

To be a speaker of English is to have acquired one of the many Slangs in a broad family, which can be roughly characterized in terms of paradigm cases and a vague intransitive notion of mutual intelligibility. This allows for graded notions of fluency/competence that let us distinguish young children, or an adult with a patchy vocabulary in a second language, from mature native speakers of a Slang family. But there is no communal Slang, Ideal English, that each speaker of English acquires yet never fully masters.

2. *Kinds of equivocality*

Of course, dictionaries are useful. They help families reduce misunderstandings. They also illustrate the difference between homophony and polysemy—cases of distinct words sounding the same, as opposed to one word having "subsenses." Polysemy may be the most interesting feature of lexical meanings, even if diagnosing examples can be hard. So let me say a little about the contrast with homophony, before turning to questions about phrasal meanings.

Words with distinct meanings can share a pronunciation. Consider 'bank', 'duck', and 'bear malice towards a bear with bare arms'. There are several English words, spelled 'bear' or 'bare', that connect their meanings with the pronunciation /bɛr/. But we don't expect *other* Slangs to connect these various meanings with a common pronunciation. Likewise, we don't expect translations of French homophones to be homophonous in English. (Consider 'seau', 'sceau', and 'saut', whose might be translated as 'bucket', 'stamp', and 'jump'.) Lexical meaning-pronunciation pairings are *arbitrary*. But 'bear a tray of food', 'bear the weight of the roof' and 'bear the pain' seem like examples of a single verb being used to talk about carrying or supporting or enduring. It's not that we have three accidentally homophonous synonyms of 'carry', 'support', and 'endure'. We gather subtly different senses of carrying, or carrying on with, under one verb.[1] But there is more than one way for a word to have multiple senses or uses.

[1] By contrast, no noun has subsenses corresponding to ursine animals and stock market pessimists; cp. 'bear a weight/wait'. It can be hard to distinguish polysemy from metaphor. So in *CM*, I don't insist on any diagnosis for any specific case. Though given any plausible way of counting lexical meanings, they allow for subsenses.

The noun 'book' can be used to talk about spatiotemporally located things that carry inscriptions of certain contents, or abstract contents that get encoded in many ways and places. It seems that 'book' lets us access dovetailing concepts that can be used to think about intimately related things that differ in ontological kind; cp. 'triangle', which can be used to talk about perceptible inscriptions or imperceptible abstracta. We can use 'window' to describe an opening in a wall, a pane of glass that occupies such an opening, a display space behind such a pane at the front of a store, an opening in an envelope that makes an address visible, or a gap in a long counter at a bank; cp. 'line', 'run', and 'set'. Even given a generous conception of homophony, a typical word exhibits a kind of equivocality, as if a word can point to a family of concepts.

If we take this idea seriously, we can view familiar examples of polysemy as special cases of a broader phenomenon: Slang lexical items are, almost always, *conceptually equivocal*. Consider the singular noun 'fish'. Following many syntacticians, I think this word combines a count morpheme with a simpler *lexical root*—often called a mass noun—that can be used to talk about the stuff in a can of tuna. The root √fish is part of the singular [ö fish+CT], whose plural form is [[√fish+CT]+PL]. I also think that *concepts* exhibit a mass/count contrast that doesn't map cleanly onto morphological complexity. The net result, I argue, is that √fish is equivocal.

Suppose that FISH_{ONE} is an atomic count-concept, WATER_{SM} is an atomic mass-concept, $[\text{FISH}_{\text{ONE}}\ \text{STUFF}_{\text{ONE/SM}}]_{\text{SM}}$ is a complex mass-concept, and $[\text{WATER}_{\text{SM}}\ \text{UNIT}_{\text{SM/ONE}}]_{\text{ONE}}$ is a complex count-concept; where subscripts on small capitals indicate conceptual types, and a slash indicates a type-converter. A child might initially link √water to WATER_{SM}, and only later introduce $[\text{WATER}_{\text{SM}}\ \text{UNIT}_{\text{SM/ONE}}]_{\text{ONE}}$ to accommodate uses of [√water+CT]. The same child might link [√fish+CT] to FISH_{ONE}, not worrying about √fish by itself until experience invites a mass concept like $[\text{FISH}_{\text{ONE}}\ \text{STUFF}_{\text{ONE/SM}}]_{\text{SM}}$. But a word for fish might also be acquired in a "stuff-first" way.

A child who is often fed fish might link √fish to FISH_{SM}, not yet realizing where fish sticks and tuna sandwiches come from; cp. √tofu and TOFU_{SM}. But such a child can still acquire FISH_{ONE}, perhaps upon hearing some swimming things described as fish, or after learning the truth about chicken(s). We can all entertain the thought that fish grows like wheat, while tofu comes from tofus that used to hop around. So we can acquire FISH_{SM} and TOFU_{ONE}, even if we already acquired $[\text{FISH}_{\text{ONE}}\ \text{STUFF}_{\text{ONE/SM}}]_{\text{SM}}$ and $[\text{TOFU}_{\text{SM}}\ \text{UNIT}_{\text{SM/ONE}}]_{\text{ONE}}$. If we know the facts, we may limit our *use* of TOFU_{ONE} to episodes of inventing stories, imagining nervous vegetarians, or considering logical possibility. But lexical roots are indifferent to the natures of what we talk about. We can use √fish to access FISH_{SM} or $[\text{FISH}_{\text{ONE}}\ \text{STUFF}_{\text{ONE/SM}}]_{\text{SM}}$; likewise for √rabbit, √chicken, etc.[2]

[2] Drawing on Brendan Gillon's work, I argue in CM that while there is no

I think this point generalizes. However words get acquired, they tend to become conceptually equivocal. But each Slang has only finitely many atomic expressions. So whatever lexical meanings are, one wants to know how they can combine to yield boundlessly many phrasal and sentential meanings. Here too, it is useful to think about ambiguity, following Chomsky.

Words can differ inaudibly because one contains a silent plural morpheme, as with 'fish'. Likewise, sentences composed of the same words can differ *structurally*. Many English Slangs connect a pronunciation of string (2) with two meanings, indicated below via (2a) and (2b).

(2) the duck is ready to eat
(2a) The duck is fit for consumption.
(2b) The duck is prepared to dine.

But while (3) has a duck-as-eaten meaning, indicated with (3a), (3) can't be used to express the duck-as-eater meaning indicated with (3b). By contrast, (4) only has a duck-as-eater meaning.

(3) the duck is easy to eat
(3a) It is easy for relevant parties to eat the duck.
(3b) #It is easy for the duck to eat relevant stuff.
(4) the duck is eager to eat
(4a) #The duck is eager to be one whom relevant parties eat.
(4b) The duck is eager to be one who eats relevant stuff.

The pattern remains the same if 'eat' is replaced with 'please' or 'love'.

The ambiguity of (2) is not due to 'ready' being homophonous. On both readings, 'ready' has its usual meaning, akin to 'suitably set, arranged, or equipped'. (Of course, 'ready' is polysemous; but so are 'easy' and 'eager'.) *Given* meanings for the word-sized pronunciations in (2), an ambiguity remains: 'the duck' can be understood as the subject of 'eat' and associated with the role of eater, or as the object of 'eat' and associated with role of thing eaten.

Similarly, whatever 'solicitor' means for you, I bet you can understand (5)

(5) a reporter phoned a solicitor from a small town

as having the meaning indicated with (5a) or (5b) *but not* the one indicated with (5c).

(5a) A reporter phoned a solicitor, and the solicitor was from a small town.
(5b) A reporter phoned a solicitor, and the phone call was from a small town.
(5c) A reporter phoned a solicitor, and the reporter was from a small town.

The attested readings reflect distinct structures: [phoned [a [solicitor

requirement that lexical roots access mass concepts, count nouns cannot be used to access mass concepts.

[from a small town]]]]; and [[phoned [a solicitor]][from a small town]]. But this doesn't explain why in the second case, 'from a small town' is understood as restricting *events of phoning* a solicitor, as opposed to *individuals who phoned* a solicitor. One can say that [[phoned [a solicitor]][from a small town]] is relevantly like '∃y:Solicitor(y) [PastPhoningByOf(e, x, y) & From-a-small-town(e)]'. But why *can't* it be construed like '∃y:Solicitor(y)[PastPhoningByOf(e, x, y) & From-a-small-town(x)]'? As the (5a)-reading illustrates, 'from a small town' can be heard as restricting a predicate of individuals. So what blocks a (5c)-interpretation of [[phoned [a solicitor]][from a small town]]?

In *CM*, I argue that meanings compose in ways that require phrasal meanings to be *monadic*, with the result that 'phoned a solicitor' has no variable for a phoner. The (5b) reading corresponds to '∃y:Solicitor(y) [PastPhoningOf(e, y) & From-a-small-town(e)]'.[3] On this view, the grammatical subject of (5) is an argument of a covert verb that combines with 'phone'.

Details aside, Slangs allow for *structural homophony* in ways that are unbounded yet tightly constrained. One string of words can correspond to two or more expression meanings, each reflecting a certain way in which the words can be arranged. To characterize these meanings, we need to distinguish the compositional character of phrasal meanings—which seems to be unlearned and common across Slangs, at least to a first approximation—from the arbitrary and often conventionalized character of lexical pronunciation-meaning (π-μ) pairs.

We also need to think about what Slangs are. Following Chomsky, I argue that they are biologically implemented *procedures that generate* certain π-μ pairs. This involves arguing against Lewis and others who think the goal is to describe *sets* of π-μ pairs, each of which can be described in various ways by the members of a community who jointly "select" the set by adhering to certain conventions. I think this E-language perspective, with 'E' connoting 'extensional', is deeply misguided. But at a minimum, we shouldn't stipulate that Slangs are sets as opposed to procedures. We should ask what Slang meanings could be such that the lexical ones exhibit arbitrary homophony and several kinds of non-arbitrary conceptual equivocality, while the phrasal ones exhibit structural homophony in unbounded but limited ways.

[3] However 'duck' is understood, 'watched her duck' corresponds to '∃y:Her-Duck(y) [PastWatchingOf(e, y)]'. Though there are twists. In (1), 'near a muddy bank' can modify 'watched'; and typically, events of watching are co-located with the watchers. So it might *seem* that 'watched her duck' has a variable for watchers. Chapter six offers more evidence that phrases are used to construct monadic concepts, and that even in 'gave a duck a dollar', 'gave a duck' does not express a relation that holds between givers and things given to a duck. I also argue that the number of arguments a verb must combine with, to form an active voice declarative sentence, often differs from the adicity of the concept lexicalized. We have polyadic concepts of eating/snacking/noshing/dining. But 'I ate' is grammatical, and it implies more than 'I ate something', while 'I snacked an apple' is not grammatical.

3. *What meanings aren't*

Even if we ignore the kinds of equivocality noted above, it seems clear that a single meaning can correspond to multiple concepts, given "Frege cases" of learning identities. Someone who learns that woodchucks are groundhogs, and that these animals are also called whistlepigs, might link at least one word to distinct concepts of the relevant rodents. But stressing Frege cases can make it seem that meanings are extensions. So I argue that a meaning can correspond to two or more concepts without being an extension that the concepts share. If meanings are "concept assembly instructions," they are *further* removed from the environment than any assembled concepts.

One might worry that my proposal conflicts with Putnam's thought experiment involving Twin Earth, where in place of H_2O there is a superficially similar though distinct substance XYZ. But if meanings are conceptually equivocal, the thought experiment is easily accommodated.

I grant that a speaker of English can use 'water' to access a kind-concept that applies—regardless of what he believes—to and only to samples of H_2O (modulo slight impurities), while his Twin-Earth counterpart uses a homophonic word to access a kind-concept that applies to and only to samples of XYZ (modulo slight impurities). But 'water' can also be used to talk about the stuff from my well in New Mexico, even though that stuff has a lower percentage of H_2O than Diet Coke, or a cup of tea on Chomsky's desk. Water from city taps often contains fluoride or worse. Ocean water is salty. This suggests that 'water' can be used to access at least one concept **C** that applies to a lot of watery stuff that is chemically *less* like pure H_2O than a lot of stuff that **C** doesn't apply to. The details seem to involve notions of sources and functional role. So absent *argument* to the contrary, why deny that 'water' can be used to access a concept that applies to the mainly-XYZ-stuff from the Twin-Earth counterpart of my well? My own *intuitions* suggest that Twin-Earthers can water their lawns, occasionally sipping water from the hose.

I grant that 'water' and 'star' can be used, in mutually comprehensible ways, by speakers who have very different views about the nature of water and stars. We can talk about stuff/things in ways that don't presuppose substantive conceptions of what we're talking about, as if words let us express kind-concepts whose contents are fixed by paradigm cases and natural dimensions of similarity. But words have many uses. And we can use 'meaning' to express a kind-concept that applies to the interpretations, whatever they are, that Slangs connect with pronunciations.

Externalism about conceptual *contents* is compatible with Slangs being procedures that pair pronunciations with recipes for assembling *concepts*. For many purposes, my Twin and I can be described as using the same recipe for how to make an apple pie (or a Negroni), even if

one of the steps calls for adding water (or ice), along with some apples (or vermouth). Recipes leave room for variation in which specific ingredients get used. Though if Earth* has only Red Delicious apples, and Earth** has only Granny Smiths, there may be contexts in which twins count as reading and following different instructions upon seeing 'take six apples' in a recipe book. Similarly, I think my Twin often uses 'water from my well' as an instruction for how to build a concept that applies to the water from his well. Though in cases where chemistry matters, 'water' and its Twin-Earth counterpart may not count as instances of the *same* word with the *same* meaning. This should be unsurprising given 'solicitor' and 'robin'.

One can be an internalist about meanings and still say that (for many purposes) my words have the same meanings as words used by other thinkers, including my former self, who have or had different concepts. But it's a trap to assume that *meanings* are whatever good translations have in common. It's even dangerous to assume that meanings are what expressions with the same meaning have in common, since 'same meaning' can be a variant of 'good translation'.

My word 'water' may count as having the same meaning as your homophonic word because each of us could add, to our own 'water'-y address, the concepts accessed via the other 'water'-y address. Likewise for 'fish', 'tofu', 'rabbit', 'democracy', etc. Much more needs to be said about *how* we exploit conceptual equivocality (and kind-concepts) to deal with the fact that we don't always think about things/stuff in the same ways. But we shouldn't assume that "sharing a language" ensures agreement about the truth conditions of sentences, as opposed to convergence on recipes for how to build thoughts. In chapters 3-5 of *CM*, I argue that Slang sentences don't have truth conditions, much less truth conditions that are determined by meanings.

Many pieces of this argument are unoriginal. Chapter three reviews Frege's contributions to the study of logic, Tarski's techniques for providing (consistent) truth theories for first-order fragments of Frege's *Begriffsschrift*, and extensions of these techniques via the use of Church's lambda calculus. I then discuss, in chapters four and five, two major difficulties for the Davidsonian Conjecture that a suitably formulated theory of truth for a Slang can serve as the core of an adequate theory of meaning for that language. First, it's hard to see how there can be true theories of truth for Slangs given examples like (6), which is my favorite sentence.

(6) My favorite sentence is not true.

Second, nonsynonymous sentences can be truth-conditionally equivalent. So it's hard to see how any truth theory for a Slang could do double duty as a good theory of meaning.

By itself, neither difficulty is fatal for the Davidsonian Conjecture. But I argue that the best hope for replying to each is at odds with the best hope of replying to the other. In any case, we shouldn't assume that Slang sentences have truth conditions, especially not if this im-

plies that (6) is true if and only if it isn't. And we shouldn't assume that a truth theory for a Slang can also be a plausible theory of understanding, given familiar objections pressed by John Foster and others. Moreover, I argue, each of these assumptions makes the other even less plausible.

Examples like (7) also tell against the idea that Slang sentences have truth conditions.

(7) Alvin chased Theodore around the tree gleefully.

Such examples illustrate the need for "event analyses," given that (7) implies both (8) and (9),

(8) Alvin chased Theodore around the tree.
(9) Alvin chased Theodore gleefully.

whose conjunction doesn't imply (7). The good idea was that (7-9) are understood as existential closures of conjunctive predicates: ∃e[PastChaseOfTheodoreByAlvin(e) & AroundTheTree(e) & Gleeful(e)]; ∃e[PastChaseOfTheodoreByAlvin(e) & AroundTheTree(e)]; and ∃e[PastChaseOfTheodoreByAlvin(e) & Gleeful(e)]. This was supposed to *support* the Davidsonian Conjecture. But a conjunct-reduction account of the implications doesn't require that (7-9) have *truth* conditions, much less that for each of these sentences, it is true if and only if some event satisfies the corresponding conjunction of predicates. In fact, requiring this leads to trouble, as sentences like (10) reveal.

(10) Theodore chased Alvin around the tree gleelessly.

Both (7) and (10) might be used, correctly, to describe a *single* episode of two chipmunks running around a tree. Alvin may have been happily chasing Theodore, who was unhappily chasing Alvin, while neither chipmunk realized who was chasing him.

There are many potential replies, involving tendentious claims about events and/or adverbial modification. But I argue that none of these replies is plausible given a moderately varied diet of examples. One important point is that while the grammatical Subject/Object asymmetry can be used to represent an Agent/Patient asymmetry, the major participants in an event of chasing (following, marrying, etc.) can be equally agentive. Related objections to the Davidsonian Conjecture can be illustrated with examples like (11).

(11) Today in London, the sun rose in the east, cars collided, and the sky was blue.

Even ignoring the polysemy of 'London', which can be used to talk about a movable polis or an immovable place, one wants to know what entities need to be posited in order to provide a plausible theory of truth for a Slang with words like 'rose', 'east', 'collide', 'sky', etc. If we agree to bracket these concerns, then alleged parade cases of characterizing meaning in terms of truth need to survive a little scrutiny.

One can say that any particular example introduces special complications. But in my view, the Davidsonian Conjecture has turned out

to be promissory notes all the way down. Attention to details keeps revealing the need for more ancillary assumptions—in part because there is a deep tension between construing event analyses as aspects of a *truth* theory, concerning how some Slang is related to what really happened, and construing such analyses as aspects of a *psychological* theory of how speakers understand expressions. Put another way, the Conjecture makes some facts about action reports look like metaphysical puzzles, and it makes some facts about truth look like puzzles about how sentences are understood. I conclude that Slang sentences don't have truth conditions, and that the Davidsonian Conjecture was fruitful but false.

4. *What meanings are (maybe)*

Identifying meanings with "concept assembly instructions" is compatible with many proposals about the relevant combinatorial operations, which determine the possible types of inputs to those operations. Indeed, the number of meaning types can range from *one*—as in a Tarskian semantics that only assigns significance to sentential expressions, and always assigns satisfaction conditions—to *endlessly many*, as in a Frege-Church semantics that invokes two "basic" types <e> and <**t**>, along with the further types licensed by the recursive principle (R);

(R) if <α> and <β> are types, so is <α, β>

where expressions of the basic types denote entities or truth values, and an expression of the type <α, β> denotes (or has as its "semantic value") a function from things denoted by expressions of type <α> to things denoted by expressions of type <β>.[4]

In my view, Slangs are not Frege-Churchy in this respect. A few iterations of (R) generates millions of types that Slangs abhor, including some that are instantiated by concepts we can easily form (e.g., concepts of the "ancestral" relation that the predecessor-relation bears to the more inclusive relation of preceding). In chapters six and seven, I also argue that phrases and proper nouns are predicative—in Frege-Church terms, instances of type <e, **t**>—and that there is little if any independent evidence for Slang expressions of type <e> or <**t**>.

I grant that Slangs can be used to build concepts that have non-predicative constituents. But these constituents may be uniformly *dyadic* and accessed by *lexical* items. As I show, this restrictive hypothesis is permissive enough to handle a wide range of constructions, including those covered by a typical first course in semantics. I posit several combinatorial operations for concepts and meanings, but only two meaning types: <M> for monadic, <D> for dyadic.

The simplest operation, "M-junction," conjoins two monadic concepts to form a third. For example, M-joining BOTTLE(_) with GREEN(_)

[4] Let's not worry here about the difference, highlighted in CM, between denoters and "unsaturated" representations.

yields a concept of green bottles; and I claim that the meaning of 'green bottle' is an instruction that is executed by M-joining concepts accessed via 'bottle' and 'green'. A second operation, "D-junction," combines a dyadic concept D with a monadic concept M to form a monadic concept of things that bear the relation expressed by D to *something* that has the property expressed by M. For example, D-joining IN(_, _) with BOTTLE(_) yields a concept of things in a bottle; D-joining AGENT(_, _) with REPORTER(_) yields a concept of things done by a reporter. I think "complete" sentences correspond to "polarized" concepts that apply to everything or nothing. The idea is that a monadic concept, perhaps assembled by executing a phrasal meaning, can be used to form a propositional concept—much as the open sentence 'Mx' can be combined with a Tarskian prefix to form the closed sentence '∃xMx', which is satisfied by all sequences (of domain entities) or none, even if 'Mx' is satisfied by some sequences but not all. I also posit a limited form of abstraction on polarized concepts; cp. 'λy.∃x(Dxy & My)'. But these operations are severely type-restricted. The resulting system is much less powerful—and much better suited to explaining *absences* of unattested expressions/readings—than familiar proposals that characterize meanings in terms of entities, truth values, functions, function-application, and a hierarchy of types.

On this view, meanings have *execution* conditions, and endlessly many Slang expressions have meanings that (unlike sets or truth values) have simpler meanings as *parts*. One can say that sentences have semantic values that are determined by values of the constituent words, given the relevant grammatical structure. But this determination thesis is, at best, an anemic explanandum. I think meanings are more like directions for how to build IKEA furniture: use a connector from box 1 to fasten a widget from box 4 to a gizmo from box 8; fasten the resulting unit to something from box 5 by using a connector from box 2; cover the result with a cap from box 9; etc. I deny that meanings are language-independent extensions of concepts that get associated with Slang expressions. I think these expressions *are* pronounceable (grammatically structured) instructions for how to build concepts; cp. perceptible (diagrammatic) instructions for how to build desks.

The last hundred pages of *CM* addresses many details concerning variables, assignments, plurality, quantifier raising, the "conservativity" of Slang determiners, and the second-order character of the concepts I appeal to. But let me end this précis by stressing that just as a lexical meaning need not correspond to a single concept, the concept *lexicalized* may not be *accessible* for purposes of composition (with other lexically accessible concepts) via relevant operations.

On any plausible view, lexicalizing a concept **C**—linking it to a pronunciation and creating a corresponding expression that has a meaning—can involve using **C** to *introduce* a formally distinct concept **C***. Given a concept that applies to ordered pairs <x, y> such that x pre-

cedes y, it might be used to introduce (i) a concept that applies to ordered triples that include truth values, or (ii) a higher-order concept that can combine with concepts like EVERYTHING and SOMETHING, or (iii) a concept that applies to events of one thing preceding another. Lexicalization need not be a mere process of labeling. Frege showed us how to introduce concepts of some logically interesting types. I think children, driven by a boring natural logic that is geared to predicates and predicate reduction, use Slangs to introduce a stock of lexically accessible concepts that are systematically combinable but *much less varied* than the diverse concepts that get lexicalized.

If this is correct, then (i) meanings play a large role in how humans acquire and combine the concepts we express with words, but (ii) focusing on truth/reference/communication is a distraction if we want to find out what meanings are. Public uses of Slangs are obvious, and they are often valuable. Though we shouldn't assume that Slangs are "for" communicating truths. Pronunciation may have been a noisy addition to procedures that generate recipes for how to build concepts. Such recipes can be useful, and sometimes worth sharing, even if they don't have truth-theoretic properties.

Croatian Journal of Philosophy
Vol. XX, No. 60, 2020

Conjoining and the Weak/Strong Quantifier Distinction

JOHN COLLINS
University of East Anglia, Norwich, United Kingdom

Pietroski's model of semantic composition is introduced and compared to the standard type hierarchy. Particular focus is then given to Pietroski's account of quantification. The question is raised of how the model might account for the weak/strong distinction in natural language quantification. A number of options are addressed and one proposal is tentatively recommended.

Keywords: Compositionality, Pietroski, type theory, weak/strong quantification.

1. *Introduction*

Pietroski (2018) offers a sparse theory of natural language semantic composition. According to the theory, the meanings of lexical items and syntactic structures are 'instructions' to fetch concepts, and such concepts are combinable by just two operations that allow for the construction of denumerably many complex concepts ('sentence meanings'). Pietroski does not affect to have solved all problems, or even to have offered a complete framework. His goal, at least as I read him, is to show how semantics might be done in a new way unburdened by many of the assumptions of the truth-conditional tradition, both in terms of compositional technology and philosophical baggage. I applaud the endeavour. The open question is how well the framework serves to capture the gamut of phenomena traditional theories target, and whether it offers novel insight. As a methodological precept, I think it is invidious to hold new frameworks to higher standards than we ask traditional ones to meet. In other words, one should always ask whether the traditional accounts are really so successful according to whatever objective standards are appropriate, for standards of explanation are often fixed by the received theoretical framework at issue. Such considerations

are particularly germane to the present case, as we shall see. Traditional semantics marries a highly expressive meta-theory with mostly descriptive desiderata. Its success at descriptive coverage, therefore, flows from the expressive resources that stand in need of explanatory justification. It is to Pietroski's great credit that he seeks to show how explanatory traction can be achieved by such minimal resources. These morals will be given some substance in the second section.[1] Thereafter, my concern will be for how Pietroski's framework might explain the weak/strong distinction in the domain of quantification. I have no settled answer here, but do examine a few options. At any rate, the distinction is an interesting topic for further inquiry into the framework on offer.

2. *Sparseness*

Pietroski's model is sparse because it eschews the 'full Frege' of semantic types:

(FF) (i) *e* and *t* are types.
(ii) If ⌜<*x*>⌝ and ⌜<*y*>⌝ are types, then ⌜<*x*, *y*>⌝ is a type.
(iii) These are all the types.

This gives us denumerably many types to map onto lexical items and their composition into phrases:

(T) (i) <e, t>: monadic predicates, such as predicative adjectives, relative clauses, and intransitive verbs (*red*, *sleep*, etc.)
(ii) <e <e, t>>: dyadic relations, such as transitive verbs (*loves*, *kicks*).
(iii) <<e, t>, <<e, t>. t>>: 2-place determiners (*every*, *most*, etc.)
(iv) <<e, t>., t>: determiner phrases (*every man*, etc.).
(v) <<e, t>, <e, t>>: attributive adjectives (*red* in *red car*, etc.)
(vi) <t,<t, ,t>>: dyadic sentence connectives (*and*, *or*, etc.)
(vii) <t, t>: sentence-level adverbs (*necessarily*, *possibly*, etc.)
(viii) <<e>, <<e, t>, <e, t>>>: prepositions, understood as VP modifiers (*about*, *to*, etc.)

In place of (FF), Pietroski presents just two compositional types, with primitive lexical content being virtually wholly monadic, and syntactic structure introducing restricted dyadic relations. Crucially, the model does *not* involve denumerably many non-applicable types as does the full Frege, i.e., types that do not correspond to any linguistic structure. The model does help itself to the 'full Chomsky' of syntactic structures, but that is independently required, unlike the Fregean hierarchy.

[1] For further discussion of these broad issues, see Collins (2020).

3. *Two types*

The first operation conjoins two predicates (simple or complex) and identifies a shared single argument position:

(M-Join) Φ(_)^Ψ(_)

Don't think of the gaps here as variables, but simply as a way of specifying the adicity of both the constituent and complex types, where the blanks are read as identical (having a co-application). Thus, it does no harm to render the result of M-Join as 'Φ^Ψ(_)'. For example, (1a) has the expected interpretation:

(1) a BROWN(_)^COW(_)
b 'BROWN(_)^COW(_)' applies to *e* iff *e* is brown and a cow.

The second operation introduces a dash of dyadicity:

(D-Join) ∃[Δ(..., _)^Ψ(_)][2]

The dyadic concepts correspond to, for example, prepositions (*above*, *with*, etc.), and also, thematic concepts. So,

(2) '∃[PATIENT(..., _)^A-BROWN(_)^COW(_)]' applies to *e* iff *e* involves a brown cow being affected.

The theory does not tell us that *e* ranges over events, as if we had an independent understanding of what events are; rather, semantically, we simply take such a complex concept as applying to 'things' that can have participants being affected, where such particpants are the kind of things that can the applicands of monadic concepts that we can M-Join to the results of D-Join.

(3) CHASE(...)^∃[PATIENT(..., _)^[A-BROWN(_)^COW(_)]]

This gives us the content for the VP *chase a brown cow* (forget about the determiner, which complicates the presentation for my purposes).

(3) above gives way to

(4) CHASE(...)^∃[INTERNAL(..., _)^A-BROWN(_)^COW(_)]

That *a brown cow* is construed thematically as PATIENT is a property determined by the verb *chase*, not the very syntax or the mere labels <V, N>. A similar story can be readily told for external arguments provided by a functional head *v* that projects to *v*P by taking a VP as a complement and a DP(/NP) as its SPEC, which is the external argument position. Thus [$_{vP}$ *v* [$_{VP}$ V N]] fetches a dyadic concept that M-Joins with whatever concept the external argument fetches (A-DOG(_), say). Thus:

(5) ∃[EXTERNAL(…,_)^A-DOG(_)]^[CHASE(…)^∃[INTERNAL (..., _)^A-BROWN(_)^COW(_)]]

[2] D-Join is not to be confused with Kratzer's (1996) *event identification* rule. The latter introduces an external argument of a verb as an agent participant of an event already specified. Thus, Kratzer's rule, like Pietroski's, involves the co-identification of a position in two composed predicates/functions, but that is where the similarity ends. D-Join specifies the character of an internal argument and introduces a further position without any thematic specification of it.

Generalising, the tactic is to treat the syntax of a phrase as fetching what we might think of as a functional dyadic concept expressed by the meaning of the lexical item as syntactically projected that provides the applicand for the open argument of the predicate (Pietroski calls such concepts *adapter concepts*).

With so much in place, let's now turn to quantification.

4. *Quantification: the generalised view*

The problem with first-order quantificational theory (as a model of NL) is that it (i) doesn't generalise across all determiners (Dets) (*most*, *few*, etc.); (ii) is wedded to an invented syntax+composition; and (iii) fails to express generalisations across Dets and within the classes of Dets.

The basic fact about first-order quantification is that it depicts natural language Dets as *sortally reducible*:

(SR) Where R is a Boolean relation and U is the universe, Q[A, B] $\leftrightarrow$ Q_U[R(A, B)]

Here we take a determiner to express a quantifier as a relation over the Cartesian product of U that specifies a cardinality for the pairs. This general approach is nowadays referred to as generalised quantifier theory (see Peters and Westerståhl, 2006, for extensive overview). SR holds for *every*, *some*, *no*, but not for *most* and other comparative relations that cannot be rendered as relations over the whole of the universe.[3] All Det relations, however, are specifiable as functions defined over the Cartesian products:

(6) a *every*: f: $< A \times B> \Rightarrow |A - B| = \emptyset$ (i.e., $A \subseteq B$)
b *some*, a: f: $< A \times B> \Rightarrow |A \cap B| \neq \emptyset$
c *no*: f: $< A \times B> \Rightarrow |A \cap B| = \emptyset$
d *most*: f: $< A \times B> \Rightarrow |A \cap B| > |A - B|$

A striking generalisation that issues from this approach is that all quantifier relations expressed by natural language Dets are *conservative* (Barwise and Cooper 1981):

(CONS) Q[A, B] $\Leftrightarrow$ Q[A, A $\cap$ B]

The truth of Q[A, B] 'lives on' the restriction A in the sense that how things are with the As alone determines truth value.

(7) a Some boy is a thief iff Some boy is a boy who is a thief
b Every girl is a swimmer iff every girl is a girl who is a swimmer
c Most women sing iff most women are women who sing

[3] In simple terms, a sentence such as *Most boys swim* cannot be rendered as a claim about the whole universe along the lines of *Most things are such that…* Although this expressive limitation of first-order quantification is widely recognised, its full philosophical consequences have yet to be properly registered. For example, deflationary approaches to truth often assume that the truth predicate is a device for generalisation over instances of a first-order scheme, but no such account can generalise to *Most things Bill says are true* (see Collins 2010).

Any theory of natural language Dets, therefore, should at least capture conservativity.

5. *The Pietroski view of determiners*

5.1. *First pass*

Take Dets to apply to ordered pairs, per the GQ approach, and to M-Join with internal arguments formed via D-Join:

(8) SOME(…)^∃[INTERNAL(..., _)^MAX:SPY(_)],

where MAX is a concept expressing the maximisation of the concept to which it applies. A predicate is formed as expected:

(9) ∃[EXTERNAL(..., _)^MAX:GERMAN(_)]

M-Joining the two, we have

(10) [SOME(…)^∃[INTERNAL(..., _)^MAX:SPY(_)]]
^∃[EXTERNAL(..., _)^MAX:GERMAN(_)]

This applies to all pairs that are such that the internal participant is an external participant, and the former is a spy and the latter is German. *Polarising*, a la Tarski, (10) applies to each pair so long as *at least one* pair satisfies the conditions (*mutatis mutandis* for other Dets).

So far so good, but the scoping behaviour of Dets is eldided.

5.2. *Second pass: QR-ed Dets*

Assume that DPs undergo syntactic movement in order to acquire scope, creating structures akin to open sentences:

(11) $[_{DP}$ Every girl$]_1$ $[_{XP}$ 1 likes Sam]

What concepts do the 'open sentences' map onto (fetch)?

Assume a concept TARSKI (Pietroski 2018: 321).

(12) TARSKI[i, P] = $(\exists\alpha)[((\forall\alpha')[\alpha' \approx_i \alpha \wedge$ ASSIGNED-BY-TO[_, α', i]) $\rightarrow$ SAT[α', P])) $\rightarrow$ SAT[α, P]]

This is not how Pietroski presents it, but it is equivalent. The basic idea is to understand open sentences in terms of their satisfaction relative to an index i under the standard Tarski condition.

Thus, we can have:

(13) [EVERY(…)^∃[INTERNAL(..., _)^MAX:GIRL(_)]]$_i$
^∃[EXTERNAL(..., _)^
MAX:TARSKI[i, _likes Sam]]

Again, so far, so good, but CONS is not reflected.

5.3. *Third pass*

On the standard treatment of natural language quantification offered by Heim and Kratzer (1998), which Pietroski uses as a foil, an open sentence is akin to a syncategorematic relative clause, but relative clauses

are not open sentences (Pietroski 2018: 337). Linguistically, raised DPs should merge with clause-like structures from which they serve as arguments of verbs. Rendering a Det as <<e, t>, <<e, t>. t>>> makes it as if a second-order relation between monadic properties, which then creates the type-mismatch problems, for <<e, t>, t> is not an <e>.

Let's take composition seriously, therefore: the DP both merges as an argument of a verb *and* merges with a clause (in its raised position). Thus, the internal and external arguments of a Det are fundamentally asymmetric, with the former restricting the range of the Det *a la* CONS. Pietroski implements this asymmetry via a modification on the TARSKI predicate.

(14) $\text{RESTRICTIVE-TARSKI}[i, P] = (\exists\alpha)[((\forall\alpha')[\alpha' \approx_i \alpha \wedge \text{ASSIGNED-BY-TO}[\_, \alpha', i] \rightarrow \text{SAT}[\alpha', P] \wedge \text{SAT}[\alpha', \text{EXTERNAL}[P'])] \rightarrow \text{SAT}[\alpha, P]]$

Again, this is my formulation, but the content is equivalent to Pietroski's definition.

We arrive, therefore, at a model of quantification that is compatible with the generalised quantifier framework, respects the movement of DPs to take scope, and can be conditioned to respect CONS. It bears noting that the definitions here are part of the meta-theory, not the compositional principles themselves, i.e., we define the relevant concepts in terms of satisfaction, but satisfaction is not part of the compositional analysis.

Hereafter, we shall look at the weak/strong distinction between natural language Dets and consider what resources Pietroski might have to capture the distinction.

6. *The distinction between weak and strong determiners*

Prior to Milsark (1977), a general distinction prevailed between definites and indefinites, but a deeper distinction was hand that has been the focus of much attention:

Weak: *some*, *a*, *no*, *one*, *two*, *few*, *many*, *several*,…
Strong: *every*, *all*, *the*, *most*, *Sam*, *both*, *neither*,…

Pro tem, think of the weak Dets as being existential, in some sense, whereas the strong Dets are universal, in some sense (clarity will be offered soon). Note that this distinction cross-classifies the definite/indefinite distinction; for example, numerical Dets are definite, but pattern with indefinite *a*, and indefinite *most* patterns with definite *every*. Also worth noting is some weak Dets can have strong construals, although not the reverse.[4] For our purposes, imagine that the distinction cleanly divides determiners into two classes.

[4] Strong DPs do not produce an ambiguity with individual- or stage-level predicates:

(i) a Every girl is clever/is in the garden
b Most boys wear shorts/have boarded the plane

7. *Three conditions*

The w/d distinction is grounded in three central semantic phenomena.

7.1. *Existentials*

There can be weak, but not strong, DP associates in existentials:

(16) a There is [$_{DP}$ a bee] [$_{C}$ in the room]
b There are [$_{DP}$ some cowboys] [$_{C}$ here]
c There are [$_{DP}$ few girls] [$_{C}$ still to see]
d */# There is [$_{DP}$ the man][$_{C}$ in the room]
e */# There is [$_{DP}$ every cowboy][$_{C}$ here]
f */# There are [$_{DP}$ most girls][$_{C}$ still to see]

Here I take the predicates (labelled 'c' for coda) to be non-constituents of the DPs in order to preclude a so-called 'list' reading, which is a specifically focused use that can rescue the unacceptable case.[5] There are also presentational readings, where there is locative, but, again, we are just interested in the existential readings.

7.2. *Weak Dets have symmetrical arguments (Keenan 1987, 2003)*

The condition here is slef-explanatory, but to be precise:

(SYM) $Q(A, B) \leftrightarrow Q(B, A)$

(17) a Some men are nurses
b Some nurses are men
c Few women are engineers
d Few engineers are women

If the first of the pairs is true, the second is true, too, so long as the determiners are weak. Strong determiners do *not* license the entailment:

If we bracket domain restriction, these sentences are unambiguously universal claims about the set of girls/boys. Similarly, weak DPs are *typically* uniform in construal across the two sorts of predicate. Obviously, the construal differs in being *existential*, in the sense in which some things are *said* to satisfy the restriction (i.e., exist), in a way (i) does not. Some determiners are atypical, such as *few*. Consider:

(ii)a Few girls are clever
b Few girls are in the garden

(iia) can only be construed as true where most girls are not clever. It has no partitive reading where some small number of girls are clever, but the rest might be smart. In contrast, (iib) precisely has such a duality of construal. It might be a claim that the garden contains a small number of girls (three, say), or the 'strong' claim that few of the girls (i.e., a small percentage) are in the garden, which might be a huge number, depending on the number of girls. This is said to be the 'strong' reading because it is about the set of the restriction as a whole, rather than some definite number of girls.

[5] A 'list' reading goes with a focused unit in response to a question, say. For example: A: Who will save us now the cavalry have left? B: Well, there is every cowboy still here. The weak cases can be expressed with the coda as a conjunct (*There is a bee and it is in the garden*) or relative clause (*There is a bee, which is in the garden*). List readings aren't so supported. The DP+codas in list readings might thus be small clauses.

(18) a Every man is a nurse
b Every nurse is a man
c Most women are engineers
d Most engineers are women

7.3. *Weak Dets are intersective*

(INS) $Q(A, B) \leftrightarrow Q(A \cap B)$.

(INS) offers a different kind of test: if 'Det As are Bs' is weak, then it is equivalent to 'Det As, who are Bs, are Cs' where $B \subseteq C$. The truth of a statement involving an intersective DP as subject wholly depends upon the intersection of the class of things that are both A and B—one may ignore the A things that are non-B. Thus:

(19) a Some man is a nurse
b Some man, who is a nurse, is a care worker
c Two men are nurses
d Two men, who are nurses, are care workers

(20) a Every man is a nurse
b Every man, who is a nurse, is a care worker
c Most men are nurses
d Most men, who are nurses, are care workers

The pairs in (19) cannot differ in truth value. Note, in particular, that both depend upon the men existing. Thus, (19b, d) are not tautological. If there are no men, say, then (19b) can't be true, and (19d) fails to be true if there is just one man. In (20), a difference in truth value between the pairs not only can obtain, but clearly does obtain in fact. (20a) is false, whereas (20b) is a tautology; ditto for (20d).

7.4. *A significant fact*

Given the *conservativity* of natural language determiners, (SYM) and (INS) are equivalent (see Peters and Westerståhl 2006: 210-11). That is:

(CONS)+(INS) entails (SYM), and (CONS)+(SYM) entails (INS)

8. *Pietroski's options*

8.1. *Pragmatics*

A ready option is to seek to capture the w/d distinction in terms of pragmatics rather than compositional semantics. It is unclear how this tactic might be realised, notwithstanding the common thought that strong utterances, as it were, presuppose existence rather assert it. We should still want to know why such a difference is tethered to the Dets and their differential behaviour with exstentials. For example, whereas a presuppositional account might explain the felt unaccept-

ability of an empty restrictor for strong Dets, the behaviour of the weak Dets remains opaque.[6]

More generally, since CONS is not a pragmatic principle, and interacts with SYM and INS, it would be nice if all these algebraic properties fell together under the one semantic explanation; indeed, the W/S phenomena look (almost) as robust as the CONS phenomena.

8.2. *Encode SYM or INS for the W cases, and let CONS do the rest*

Recall our significant fact. CONS+SYM entails INS, and CONS+INS entails SYM. If CONS is encoded, therefore, we really only need to encode one of the other properties. Perhaps the easiest implementation of this idea is to let the internal predicate be RESTRICTED-TARSKI modified by the satisfaction of the external predicate for the weak Dets alone. Such would satisfy SYM, and so entail INS.

I can see two main problems with this thought. Firstly, it is not obvious how to make it work compositionally, i.e., why should there be a restriction going up the syntactic tree? The restriction going down tracks the syntactic movement of the DP. In short, it appears to be a stipulation. Secondly, the *there*-existential restriction to weak Dets would remain unexplained. I shall come back to this shortly below.

8.3. *An alternative*

Suppose that only the weak Dets encode or fetch an empty internal monadic concept that we may render intuitively as '*x* is in the domain', but when combined with the Det effectively encodes the idea that a cardinality of things that satisfy the predicates exists. This captures the existential content of the weak Dets and simultaneously explains why the strong Dets admit empty restrictors. For example:

(21) [SOME(…)^DOMAIN(…)^ (∃[INTERNAL(..., _)^MAX:GIRL(_)]$_i$ ^∃[EXTERNAL(..., _)^MAX-RESTRICTED-TARSKI[i, _likes Sam]]

Being more speculative, we may think, on this story, that weak Dets 'originate' from a means of talking about some 'relevant' domain, whereas the strong Dets don't. Thus, the latter don't introduce a domain but have a global or universal meaning (more anon).

Let's see how this basic idea might fair in accommodating the basic properties of the w/s distinction.

Firstly, the proposal accommodates SYM, for with weak Dets, a domain is populated with a kind of thing (or things) that has the internal property, which is also said to have the external property, and so

[6] It is common to think of strong Dets as presupposing a non-empty restriction, hence the supposed infelicity of *Every French king is bald*. It is better, I think, to account for the infelicity as due to implicature.

whatever condition the Det places on the thing(s) having the internal property will hold of the external property too. Not so for strong Dets, where no domain is populated by things.

Secondly, the proposal accommodates INS, for, again, whatever kinds of things are in the domain share the properties of the internal and external predicates. With strong Dets, since there is no domain, we can have an empty-restrictor reading in line with the non-intersective relation of the strong Det.

Thirdly, the proposal also accommodates the fact that some weak Dets can be read strongly. For example, the weak construal of *few* is as expected, with the domain populated by few things that are both girls and in the park, as might be. On the strong reading, the domain is still populated (no empty-restrictor reading is available), but some broader group of girls than just those in the park must be understood to exist. In effect, the content is partitive.

Fourthly, the *there*-existential restriction to weak associates is nicely accommodated. In the weak cases the domain is populated, which is actually just what the bare existential says. The strong Dets encode no domain, and so they have no existential reading, unless a domain is explicitly introduced via presentation or 'list', which are, of course, supported by strong Dets.

In the following section I shall dwell somewhat on existentials, for they add some interesting support to my general proposal.

9: *Some syntactic considerations (after Kayne 2019)*[7]

According to Kayne (2019) there are four possible construals of *there* exhibited in (22):

(22) **There**$_{e}$ are two files **there**$_{l}$ on the desk, which **therefore**$_{r}$ need filing in them **there**$_{p}$ cabinets.

So, there can can be existential, locative, rationale, and presentational readings. Instead of positing a 4-way ambiguity, let *there* have a basic presentational construal fixed in a low small clause with its associate; the other construals are configurationally fixed. In particular, for the existential, *there* obligatorily moves to SPEC-TP (subject). Of course, this makes sense of the general syntactic differences, between existential *there* and locative *there*.[8]

[7] Herburger (2000) offers other syntactic reasons why strong Dets must move to SPEC-TP

[8] Neither of the them receive a theta-role, but locative *there* is not an argument, while existential *there* is, albeit an expletive. Syntactically, this shows up in various ways. Firstly, locative *there* cannot occur in a tag question:

(i) *There is your dog, isn't there?

Secondly, it does not admit raising:

(ii)a There seems to be a dog in the garden

b *There seems to be your dog

Thirdly, locative *there* cannot be negated:

(iii) *There isn't your dog

None of this by itself essentially bears on the w/s distinction. Kayne merely notes that there must be 'some constraint'. Note, however, that the existential *there* is still presentational content-wise; it is only existential configurationally. Suppose, then, that *there* is always first merged in a small clause with its associate and interpretation happens throughout the derivation, not just at LF or some other completed structure. If we now also posit a domain predicate with weak Dets, then the small clause will be interpretable with presentational *there*. If the Det is strong, and so lacks a domain, the presentational *there* will be uninterpretable. Thus, we get to explain the existential restriction on the assumption of Kayne's model and the weak Dets introducing a domain predicate.

10. *Conclusion*

I hardly think what I have said here is the end of the matter, or even the beginning of the end of the matter. I suspect it might not even be the end of the beginning of the matter. I commend the topic for further inquiry within Pietroski's basic framework.[9]

References

Collins J. 2010. "Compendious assertion and natural language (generalized) quantification: a problem for deflationary truth." In C. Wright and N. Pedersen (eds.). *New Waves in Truth*. London: Palgrave Macmillan: 81–96.

Collins, J. 2020. "Conjoining meanings without losing our heads." *Mind and Language* 35: 224–236.

Heim, I. and Kratzer, A. 1998. *Semantics in Generative Grammar*. Oxford: Blackwell.

Herburger, E. 2000: *What Counts: Focus and Quantification*. Cambridge: MIT Press.

Kayne, R. 2019. "The unicity of *there* and the definiteness effect." In *Questions of Syntax*. Oxford: Oxford University Press: 134–166.

Keenan, E. 1987. "A semantic definition of 'indefinite NP'". In E. J. Reuland and A. ter Meulen (eds.). *The Representation of (In)definiteness*. Cambridge: MIT Press: 286–318.

Keenan, E. 1993. "Natural Language, Sortal Reducibility and Generalized Quantifiers." *Journal of Symbolic Logic* 58: 314–325.

Keenan, E. 2003. "The definiteness effect: semantics or pragmatics?" *Natural Language Semantics* 11: 187–216.

Kratzer A. 1996. "Severing the External Argument from its Verb." In J.

Fourthly, locative *there* cannot occur in a subordinate clause:
(iv) *Sam wondered whether there was your dog.

[9] My thanks go to Paul Pietroski for conversations on the topic of this essay and many related issues, and to Michael Glanzberg for the same. I was hoping to find time to develop a more worked out discussion, but children, pandemic, and semantics are not a happy mix.

Rooryck and L. Zaring (eds.). *Phrase Structure and the Lexicon*. Dordrecht: Springer: 109–139.

Milsark, G. 1977. "Toward an explanation of certain peculiarities of the existential construction in English. *Linguistic Analysis* 3: 1–29.

Pietroski, P. 2018. *Conjoining Meanings: Semantics without Truth Values*. Oxford: Oxford University Press.

Peters, S. and Westerståhl, D. 2006. *Quantifiers in Language and Logic*. Oxford: Oxford University Press.

Croatian Journal of Philosophy
Vol. XX, No. 60, 2020

Compositionality and Expressive Power: Comments on Pietroski

ELMAR UNNSTEINSSON*
University College Dublin, Dublin, Ireland
University of Iceland, Reykjavík, Iceland

Paul Pietroski has developed a powerful minimalist and internalist alternative to standard compositional semantics, where meanings are identified with instructions to fetch or assemble human concepts in specific ways. In particular, there appears to be no need for Fregean Function Application, as natural language composition only involves processes of combining monadic or dyadic concepts, and Pietroski's theory can then, allegedly, avoid both singular reference and truth conditions. He also has a negative agenda, purporting to show, roughly, that the vocabulary of standard truth conditional semantics is far too powerful to plausibly describe the linguistic competence of mere human minds. In this paper, I explain some of the basics of Pietroski's compositional semantics and argue that his major objection to standard compositionality is inconclusive, because a similar argument can be mounted against his own minimalist theory. I argue that we need a clear distinction between the language of the theorist—theoretical notation—and the language whose nature we are trying to explain. The theoretical notation should in fact be as expressively powerful as possible. It does not follow that the notation cannot be used to explain mere human linguistic competence, even if human minds are limited in various ways.

Keywords: Compositionality, semantics, minimalism, internalism, Paul Pietroski, function application.

* Many thanks to the participants and organizers of the *Philosophy of Language and Linguistics Conference* in Dubrovnik 2019, part of which was devoted to Pietroski's new book. Special thanks are owed to Dunja Jutronić for her help and patience. For discussions of the topics of this paper I am indebted to Paul Pietroski, John Collins, Michael Glanzberg, David Pereplyotchik, Daniel Harris, Finnur Dellsén, Una Stojnic, Ernie Lepore, Barry Smith, Anna Drożdżowicz, Michael Devitt, Paolo Santorio, and David Sosa. This work was supported by the Icelandic Research Fund (grant no. 163132).

1. *Introduction*

In his book, *Conjoining Meanings: Semantics without Truth Values* (2018), Paul Pietroski develops his own minimalist compositional semantics, based on only a single compositional principle which takes every complex expression to encode a monadic concept. This is hailed as a much more plausible theory of semantic competence than alternative views that trade in an infinite number of expression types and, usually, more than one principle of composition. The result is a robust and serious internalist alternative to the existing externalist orthodoxy, which Pietroski believes is too mired in the vocabulary of extensions, functions, and truth conditions.

In this paper, I offer a very rough sketch of Pietroski's positive proposal, explaining what he takes meanings to be – namely, instructions – and how they compose (Section 2). Next, in Section 3, I present his argument for thinking that standard truth conditional semantics is far too powerful to be an appropriate tool for describing semantic competence in normal human beings. I also describe one of his arguments against function application as a compositional principle, namely that it misrepresents monadic concepts as relations. According to Pietroski, the concept $F(x)$ is not a function from objects to truth values – making a monadic concept relational – but a mental device of classification, to classify things as F. I argue, however, that the two proposals are either incomparable, or, if they are made comparable, they may just as well turn out to be ontologically and theoretically equivalent. This all depends on further commitments, not encoded in the mere notation for function application, which remain optional for the truth conditional semanticist.

In Section 4, however, I argue that Pietroski's own semantics would be subject to objections very similar to the ones he presents against truth conditional semantics. Briefly, if we think of his own proposal in terms of basic syntactic types and some function with those types as its domain, it is easy to see how his proposal will generate a boundless number of new types. If this is right, Pietroski himself seems to assume compositional capacities that are, by his lights, too powerful to be ascribed to finite human minds. I conclude, on the contrary, that this shows that the argument itself is flawed. Roughly, the strength of the notation used by the theorist is no guide to the actual metaphysical commitments assumed in the theory itself. So, as before, I have failed to find a substantial cause for disagreement between Pietroski and his alleged opponent.

Finally, in Section 5, I describe a distinction that might be of help in this debate. This is the distinction between the representational system that the theorist uses to explain and describe some cognitive phenomenon and the system employed by the cognitive agent or agents under scrutiny. There are reasons to think that, in principle, these cannot be identical. If so, we are free to take truth conditional semantics, and Pi-

etroski's minimalist semantics, as competing models of the same thing, rather than, say, attempted reproductions of what actually happens in human minds. And this is the natural conclusion, showing that comparing the two models may be even more difficult than Pietroski assumes.

2. *Slang and slanguages*

Shaking his head over the many meanings of 'language,' Pietroski defines 'Slangs' as naturally acquirable human languages. The languages of logic, mathematics, musical notation, and so on, are not at issue. More precisely still, Slangs are constituted by a collection of mental processes that determine sets of pronunciation-meaning pairs. Presumably, sign languages are Slangs, so 'pronunciation' must also apply to the manner of producing sign expressions. Long ago, David Lewis (1975) would have defined Slangs as the sets of pronunciation-meaning pairs themselves, but Pietroski thinks this is a mistake. Still, we are all people of the same trade; semanticists working to discover the nature of the meanings or interpretations generated by the mental processes in question.

Pietroski's enduring methodological commitment is to the mantle of compositionality. The only thing we know for certain about the semantic properties of Slang expressions is that they must, must compose. Composable Meanings (CMs) are then identified as instructions for fetching and assembling human concepts. To see the point, it is best to think about what CMs cannot be. First, these meanings are not human concepts. For example, if there is a human singular concept of the person Pierre, this concept cannot be the Composable Meaning of the Slang expression 'Pierre.' Second, CMs are not extensions. For example, if the extension of the human singular concept of Pierre is Pierre himself, this entity cannot be the CM of 'Pierre.' Third, CMs are not instructions for how to use expressions or pronunciations. For example, if the pronunciation /pierre/ ought to be used to refer to Pierre, the instructions so to use it cannot be the CM of 'Pierre.'

More positively, CM-semantics involves instructions which are both composable and executionable by human minds. There are two kinds of instructions, or mental processes, *fetch*-processes and *assemble*-processes. Fetching is simple. Take the polysemous Slang expression 'book.' The Composable Meaning of 'book' is an instruction to fetch a concept stored at the 'book'-address. Since 'book' is polysemous, this particular address is home to more than one concept. Say there are two 'book'-concepts, one of books as concrete entities we can buy or burn, another of books as abstract collections of information we can choose to forget. If so, the CM of 'book' is an instruction with two admissible executions, either to fetch the concept BOOK:CONCRETE or the concept BOOK:ABSTRACT. To understand the Slang expression 'book' on a given occasion, is to execute the associated instructions to fetch either of the two concepts stored at the appropriate lexical address.

It is worth pausing to think about how meanings are being individuated on this account. Strictly speaking, 'book' has only one meaning, because meanings are instructions and the instruction for 'book' is to fetch a concept at a given address in the mental lexicon. The polysemy of 'book' consists in the fact that the address is home to two concepts and, thus, executing the instructions on a given occasion can have either of two end results. More natural, perhaps, would be to say that meanings are constituted by sets of admissible instruction-executions. That way, 'book' would have two meanings according to CM-semantics. Let's reserve the expression 'recipes' for ordered pairs of instructions and admissible executions. Polysemous expressions are expressions with more than one recipe, that is, more than one meaning. As we will see, I suspect that Pietroski must insist that Composable Meanings are instructions and not recipes. And he might very well be right.

A few words on fetching and assembling. To fetch a concept is a specific mental process which ranges over composable human concepts and is triggered by atomic expressions like 'red' and 'dot.' The lexical addresses already mentioned can only be home to composable concepts, not any old human concept. Further, these concepts can either be monadic or dyadic, but not triadic or more. Well-formed instructions to fetch a concept can themselves be composed to yield more complex instructions. Complex instructions of this sort, encoded by molecular expressions like 'red dot,' trigger the appropriate assemble-process. There are two major processes of this sort, called M-join and D-join, the results of which are unsaturated predicative concepts. So, assemble-processes take fetchable concepts and either M-join them or D-join them. 'Red dot' is M-joined to yield the concept RED DOT(_), applying to red dots. An expression like 'above' fetches a dyadic concept, ABOVE(_, _) with two unsaturated argument places. This concept may be composed with RED DOT(_) with the D-join operation, to yield a monadic concept applying to whatever is above a red dot, roughly ABOVE RED DOT(_). Two dyadic concepts cannot be joined, only two monadic ones or a monadic and a dyadic one. And the result is always a composable monadic concept. Here I suppress a number of important details, for example about how exactly D-join works, because they are not important for the points I wish to make.

I should stress that, for CM-semantics, it is entirely possible that humans possess and regularly employ unfetchable and non-assembling natural concepts. The Composable Meaning of 'Pierre' is an instruction to fetch a composable concept at the 'Pierre'-address. According to Pietroski, we better think of this as a special monadic concept, PIERRE(_), applying to objects which are called 'Pierre' (249). This is what makes the concept composable via the assemble processes of M-join and D-join. Even so, the CM-semanticist can very well allow that hearing the expression 'Pierre' uttered on a given occasion may come to activate another concept in one's mind, namely the singular con-

cept PIERRE (108). But activation and fetching are not the same thing. Meanings may activate a number of mental phenomena, some even regularly and reliably, but, still, these mental phenomena will not be composable concepts.

Or, so the story goes at least. The Fodorians in the room might well wonder what, if anything, can be meant by the idea of a concept that does not compose. If there is a language of thought ('Mentalese'), presumably it enjoys a compositional semantics. So, if there is a singular concept of Pierre, the denoter-concept PIERRE as it were, then it must compose with other concepts, like the concept SNORES, to yield a proposition with a truth condition, being true if and only if Pierre snores. If we think in a compositional Mentalese language—which, we should remember, need not be a Slang—the theory of CM-semantics may appear inherently unstable. Put it this way, assuming that Mentalese contains the purely denotational expression 'Pierre' whose content is exhausted by the individual, Pierre, and thus, that such Mentalese expressions must compose (for Mentalese is compositional), why shouldn't Slangs be able to contain such expressions too? If Mentalese-'Pierre' can fetch singular concepts—because they are compositional—why can Slang-'Pierre' not do so as well?

Not to dwell too long on this point, I believe Pietroski would be best served either to deny that there are any singular terms in the standard Kripkean sense, or to deny that we think in a compositional language of thought. As far as I can see, both may be sources of prevarication on his part. Understandably, as both are non-negotiable for some theorists.

This is my rough sketch of Pietroski's positive proposal, leaving out a lot of fascinating detail, but he also has a more negative agenda. He argues that this minimalist theory is in tension with standard possible worlds truth conditional semantics and, further, that the latter was a bold conjecture best consigned to the flames. In the next section I will focus on one particular argument in this vein, suggesting that it is inconclusive as it stands. First, however, I think it may be helpful to state more flat-footedly what seems to be the real difference between maximalist semantics—as we might call the alternative—and Pietroski's minimalism.

The minimalist rejects function application as a basic compositional principle, putting in its place something more like predicate modification (assemble processes). Both of these principles are tentatively endorsed in the standard textbook of maximalist semantics (Heim and Kratzer 1998), where the elimination of predicate modification is perceived as desirable but ultimately undoable. Heim and Kratzer do not seriously consider the option of eliminating function application. Pietroski's contribution, among many other insights, is at least that of presenting a robust theory on which only something like predicate modification is assumed. Ultimately, however, I think one question is

left hanging in the air: Has it been shown, or even made plausible, that there would be a distinction at the level of mental mechanisms between a system with maximalist psychology and a system with a minimalist one? That is, is there any cognitively realized distinction between systems characterized by the assemble function and those characterized by function application together with predicate modification? As far as I can see, but certainly Pietroski disagrees, these may well turn out to be equivalent models of the same cognitive phenomenon, depending on the precise commitments of the maximalist theory. If so, then, as I will argue in the next section, we should stick to the more powerful theoretical vocabulary of maximalist semantics.

Finally, more sweepingly, minimalists need stronger arguments to eliminate extensions, truth conditions, truth values, and possible worlds from any semantic theory, even their own. Pietroski does not rule out that Mentalese has a maximalist semantics (84-85). But he needs to show that this is not the case for, otherwise, the argument for having one type of semantics for both Slangs and Mentalese come knocking. That's not the only problem, however, because if Slangs are used to activate Mentalese to, in turn, activate propositional attitudes, the connection between the two seems far too tight to argue that Slang-'Pierre' does not function, somehow or other, to fetch or activate whatever Mentalese-'Pierre' is supposed to fetch or activate. Minimalism can carve out semantic space which excludes extension and truth conditions, but the carving itself does not show that such phenomena do not exist or that they are irrelevant to the study of meaning and content. To carve things up in this way is merely to insist that meanings are instructions and not recipes, as these terms were defined above.

3. *The Fregean hierarchy and classification*

Pietroski indeed presents one very ambitious argument of this sort. If sound, it seems, it would establish precisely that neither Mentalese nor Slangs could possibly enjoy a maximalist semantics. The conclusion of the argument is, roughly, that the notational machinery—specifically the Fregean Hierarchy of Types and the lambda-calculus—in which maximalism is entwined, is far too powerful and sophisticated to represent the mental processes of finite human minds. In what follows, I focus on this argument, which, although impressive and intriguing, is inconclusive.

Maximalist semantics is standardly introduced by defining so-called Fregean or Montagovian semantic types (e.g., Dowty et al. 1981, Heim and Kratzer 1998) and Pietroski points out that the types are 'boundlessly many' (127). The types are defined recursively as follows:

(1) e is a type
(2) t is a type
(3) if a and b are types then $<a, b>$ is a type

This definition represents an infinite class of syntactic or semantic categories. It follows, for example, that <*e*, *t*>, <<*e*, *t*>, *t*>, and <<*t*, *t*>, <*t*, *t*>> are all Fregean types. Expressions of type *e* refer to objects like Pierre or Luang Prabang. Expressions of type *t* stand for truth-values, True or False. Slang expressions like 'snores' are recursively defined functions from objects to truth-values. So, 'snores' is a type of expression which takes (the semantic value of) an *e*-type expression and outputs a truth-value. From this we get the intuitively correct result that 'Pierre snores' is true if and only if the function encoded by '*x* snores' outputs truth when its argument place is assigned the object referred to by 'Pierre.'

Pietroski points out that we can define levels of complexity for the Hierarchy of Types. (1) and (2) in the recursive definition is Level Zero. Level One is defined as the set of possible functions <*a*, *b*> whose members *a* and *b* are at Level Zero. So, Level One is <*e*, *e*>, <*e*, *t*>, <*t*, *e*>, and <*t*, *t*>. Each higher level is defined in the same way, such that Level *N* builds types from all levels lower than *N*. Level Zero has two types, One has four, and Two has thirty-two (for example, <<*e*, *t*>, *t*>). When we reach Level Four, we have more than two million types. Many types at Levels Three and Four seem bizarre and are not realized, Pietroski states, by human minds. But we certainly can define such types, and some might be important in logic or mathematics, like the Fregean concept of an ancestral, which is Level Four (128). Pietroski wants to conclude that Slangs cannot be Fregean languages, because they are far too powerful to be realized in the mind of every human who speaks a natural language. If this follows for Slangs, it follows for Mentalese as well. Fregean thoughts are for Frege, not us mere mortals.

Maximalists semanticists also use the extremely powerful lambda-calculus to make good on their promise of a truth conditional semantics. As Pietroski shows, the Fregean idea, which required Tarski's notion of quantification over variant sequences of objects, is to treat monadicity as a special case of relationality (83). For example, the semantic value of 'brown' is a special kind of truth function defined on the basis of lambda-abstraction, λx<*x* is brown>, mapping *x* to Truth if *x* is brown and to Falsity otherwise. So, this is a function, roughly, taking one set of objects to T and another set to F, determining sets of things satisfying a certain condition, namely having the property of being brown. In principle, this allows maximalists to model any Slang expression which is not of type *e* or *t* as a mathematical function. This is implicit in the Fregean Hierarchy already discussed, where 'brown,' for example, is an expression of type <*e*, *t*>. But the lambda-calculus is necessary to make this idea coherent and workable.

To the contrary, Pietroski argues, monadic concepts like BROWN(_) do not represent functions at all, especially not a sophisticated lambda-function from objects to truth-values. Rather, he claims, such concepts are classificatory; they let us classify objects without relating them to truth values (p. 83). He explains that there is a psychological distinc-

tion between relational and classificatory concepts, the former are for '... classif[ying] things, into those that meet a certain condition (e.g., being a rabbit) and those that do not. Anything that meets the condition satisfies the predicate, which applies to anything that meets the condition.' (28). Relational concepts, however, are ones like ABOVE(_, _), where one object is related to another and the relation is satisfied when the objects are in fact, in this example, related such that one is above the other. Pietroski's point is that BROWN(_) is simply not relational and thus it is misrepresented by the Fregean theory.

But what is the difference exactly? If we tried to construct a creature whose monadic concepts are non-relational because they are merely classificatory, what would we get? It seems almost impossible to end up with anything other than a relational concept in Frege's original sense. This is because classifying things into those with the property F and those that do not have the property F is equivalent to classifying things into those that are truly judged to be F and those that are falsely judged to be F. At least, that seems to be Frege's position, as Sanford Shieh (2019) has recently tried to show in some detail. Briefly, Shieh's interpretation is that Frege believed that the truth of the thought that *p* is constituted by the obtaining of what the thought represents. To recognize that the referent of '*a*' falls under the referent of 'is *F*' is thereby to recognize that the thought that *a* is *F* refers to the True (2019: 108). If this is right, there is no difference at all between classificatory concepts and Fregean concepts with a single unsaturated argument place, indicated by '*x* is *F*' (Glanzberg 2014: 267 makes a related point I think).

Still, I don't think this objection is conclusive. What we would need is some account of the difference between classifying and relating to truth values and, in fact, I think there might be a plausible account of this sort. But it is not one I can find in Pietroski and of course that might be my own fault. Anyway, very roughly, if we hold that non-declarative clauses, like imperatives and interrogatives, do not relate objects to truth values, because they have no truth conditions, then the properties or concepts occurring in those clauses must classify objects without relating the objects to truth values. But the very same properties can occur in declarative clauses too, and, so, we have something of puzzle. As I understand most proposals in the literature on imperative semantics, for example, they ultimately model properties—even as they occur in imperatives—as functions from objects to truth values (see, e.g., Roberts 2018). I hope to address this issue in future work, so I leave it unresolved here.

As if by the law of gravity, this discussion is veering dangerously close to the whole issue of the connection between logic and psychology, and Frege's own complicated view of that connection. The topic is too massive, controversial, so we will mostly steer clear. But Pietroski

writes that the relational conception of monadic concepts, and the Function-Argument structure of thoughts more generally,

> ... led to a brilliant conception of how thoughts could be logically related. But like Frege, I don't think it should be viewed as a psychological hypothesis, even if it can be viewed as a model for a certain kind of ideal cognition. (83)

It would seem, however, that Frege's anti-psychologism is so staunch as to exclude any psychology whatsoever, no matter how powerful. His anti-psychologism is driven by the conviction that logic is an autonomous science, not reducible to (anyone's) psychology. Thoughts are timelessly and mind-independently true or false. Ditto, then, for the logical relationships between different thoughts (Frege 1918/1956).

Still, undeniably, thoughts in Frege's sense are relevant to psychology, because they are what we grasp, judge, and assert. Even more, they are what any cognitive creature, ideal or not ideal, would grasp, judge, or assert, in trying to discover true thoughts, which for Frege is the 'work of science' (Frege, 1918/1956). If ideal cognition discovers the true thought that p, and the truth of p consist in the fact that a is F, then no one else can discover exactly the same thing without recognizing that a is F. For Frege, a model of ideal cognition is a model of cognition. So, if the Function-Argument structure reveals to us (the theorists) how certain thoughts are related logically, we have also found out which particular thoughts need to be grasped by anyone credited with recognizing those relations. The structure itself would then seem to be theoretically indispensable.

Finally, the quote may help us better to diagnose the problem of distinguishing classificatory and relational concepts. As Pietroski says, there is a sense in which Frege does not intend the relational analysis as a psychological hypothesis. What this means for Frege, I take it, is that the structure imputed on the thought by the logician has no direct psychological relevance. That is, it does not tell us anything about the nature of the mental acts of grasping, judging, or asserting; except insofar as those acts are individuated in terms of their objects. More importantly, though, if the Fregean structure is not a psychological hypothesis in this sense, then there is no tension between that structure and any other structure, when it comes to human psychology. The relational conception of monadic concepts, and monadic thoughts like a is F, can very well model whatever it is that the classificatory conception models. At the level of psychological mechanism, the two are equivalent. If both range over exactly the same set of thoughts—if we are allowed the Fregean notion of a thought—they are completely equivalent.

4. *Notation and expressive power*

Pietroski subscribes to the following methodological strategy:

> In defending any proposal about meanings, one must also take care to not assume implausibly powerful expression-generating capacities. Similarly, in defending any proposal about syntactic structures, one must take care to not assume implausibly powerful operations of semantic composition. (294)

Surely, the counsel is sound. But how exactly do we judge degrees of power? And when we know how, will it really follow that Pietroski's minimalism has the right degree and maximalism not? My response to the first question is, one, that at least it is not judged by the expressive power of the notation employed by the theorist and, two, that any theory powerful enough to describe the basic range of human semantic competence, has exactly the right degree of expressive power. It may, still, lack explanatory power, which is a different matter. My response to the second question is that, no, this particular line of argument against maximalism is inconclusive, at least if both parties are working with some distinction between competence and performance.

There appear to be two notions of power in play here, productivity and composability. A theory's degree of productivity is determined by the number of new items, or syntactic types, it can generate from some finitely stateable base of principles. Generative theories of semantic competence are normally thought of as generating a potential infinity of new items. The degree of composability, however, is determined by the number of new items or types which can serve as inputs into the compositionality function postulated by the theory. Presumably, if the items are infinite in number, the domain of the compositionality function will be infinite as well.

The primary virtue of Pietroski's compositional principles is their simplicity. The procedures modeled by these functions are dumb and can certainly be performed by specialized cognitive mechanisms, rather than needing the resources of a full-blown mind. But this virtue is shared by function application.

To see this more clearly, consider the productivity and composability implicit in the notational systems we unhesitatingly, and rightly, employ in describing the thoughts of young children. We posit that young children can represent the thought that John thinks that cows are brown and that 2 + 2 = 4. In the first case, we employ the incredibly powerful formula '*A* thinks that *p*.' By simple recursion, we will get '*A* thinks that *B* thinks that *p*.' So, postulating thoughts about others' propositional attitudes involves notation which is infinitely productive and boundlessly compositional. Similarly, in the second case, we employ the concept of addition, symbolized by '+'. In particular, we attribute the thought that 2 + 2 = 4 to children well before they develop the capacities to represent numbers higher than 100. If Pietroski is right, this should be a problem. The symbol '+' is infinitely productive and boundlessly compositional and thus should not be used to describe the

thoughts of young children. But that would be the wrong conclusion to draw.

Moreover, CM-semantics can be developed recursively as a Hierarchy of Types, just like the maximalist theory (Pietroski mentions this possibility briefly on p. 113). CM-semantics has two principles of composition, M-join and D-join, both of which will have monadic composable concepts as their outputs. In M-junction, if Φ(_) and Ψ(_) are composable concepts then Φ(_)^Ψ(_) is also composable. Here '^' stands for the compositional process of conjunction, let's call it 'junction.' In D-junction, if Φ(_) and Ψ(_,_) are composable concepts then Ψ(_,_)^Φ(_) is a composable monadic concept, for example the concept for classifying things as being above a red dot. A brown cow above a red dot would be an example where the results of M-junction and D-junction are themselves joined by M-junction. Let's now define the Pietroskian Hierarchy by recursion:

(1) m is a type
(2) d is a type
(3) if a and b are types then (a^b) is a type

Here, m stands for monadic concepts and d for dyadic concepts. The peculiarities of junction ('^') are important, as it behaves very differently from the set theoretic notion of an ordered pair employed by the maximalist. With that in mind, if Level Zero of the Hierarchy is exhausted by m and d, Level One is exhausted by the two types (m^m) and (m^d). Level Two is only slightly more complex, if defined in terms of the number of added admissible junctions we can perform by the junction of types at Levels Zero and One, e.g., [m^(m^m)] and [(m^d)^(m^d)]. As soon as we reach Level Three we have 56 types, Level Four has 2,212, and Level Five has 2,595,782 types. We have clearly outstripped the finite cognitive capacities of mere human minds. Still, Pietroski is happy to employ notation which belongs to Level Four (202), if interpreted as [(d^m)^m]^[m^(d^m)]^m:

∃[AGENT(_, _)^AL(_)]^CHASE(_)^PAST(_)^∃[PATIENT(_, _)^THEO(_)]^GLEEFUL(_)

This would be a proposed logical form for 'Al chased Theo gleefully,' making explicit who is the chaser (agent) and who the chasee (patient). I have not explained Pietroski's notion of existential closure, '∃,' but it is needed to coordinate one of the gaps in the d-type with the gap in the m-type. So, roughly, x is above a red dot if there is a red dot y such that x is above y. Notice, also, that in CM-semantics, juncture-processes are insensitive to order, so [(d^m)^m] is equivalent to [m^(m^d)].

Why does Pietroski not see this as a problem for his minimalism? That is, why does he assume that this form of argument is a blow to maximalism and not to his own theory? I think there are two reasons which, although illuminating, are inconclusive.

First, in considering something like this point, Pietroski seems to assume that (3) is not really true. That is to say, he believes CM-

semantics has only two syntactic types, *m* and *d*, and claims that all complex expressions would be of type *m*. Therefore, there are no non-basic types in his semantics (113). I do not see how this could be true. Junction is a process or operation modelled by a function. The *value* or *output* of the function is always an expression or concept of type *m*, but that doesn't mean that junction itself is of type *m*. That would be like saying that <*e*, *t*>-expressions are really *t*-type expressions, because their output is always a truth-value. Rather, junctions are functional expressions, whose domain could be defined as follows. Take the set *D* consisting of (i) all items of type *m*, (ii) all sets {*m*, *d*} of items of type *m* and *d*, and (iii) the infinite number of sets resulting from every possible combination and repetition of the elements in (i) and (ii). Now remove every non-set from *D*, that is, delete the element *m*. Next, take the set *G* of all items of type *d* and define the Cartesian product of sets *D* and *G*, $D \times G$ (so we get, for example, {{*m*, *d*}, *d*}). Call the resulting set *F*. At last, the junction-function is defined as a mapping from every element in *F* to a set of expressions or concepts of type *m*. For example, this function takes the pair BROWN(_) and COW(_) as input and yields the monadic concept BROWN(_)^COW(_), which classifies things as brown cows. *F* contains infinite sets, so it goes without saying that human minds are unable to apply junction to every member of *F*.

Similarly, maximalist semantics really contains three types, *e*, *t*, and *f*, where *f* is a function defined in terms of the first two types. CM-semantics contains expressions of type *m*, *d*, and *j*, where *j* is a function from *F* to a set of *m*-types. This way of looking at the matter is unavoidable, because otherwise we will lack the resources to distinguish the composed concepts Φ(_)^Ψ(_) from the non-composed composable concepts Φ(_). Both are indeed composable, but only the former is composed with the operation of junction. We can think of *j* as a function from, for example, {{*m*, *d*}, *m*} to [(*m*^*d*)^*m*], where the latter is itself a monadic concept of a certain sort. If this is right, junction is both infinitely productive and infinitely compositional. To be clear, I do not hold that this is a problem for CM-semantics, to the contrary. My point is only to argue that if too much notational power is a problem for maximalism, for the reasons Pietroski adduces, it is also a problem for minimalism. The relative power of the notation is irrelevant to the basic metaphysical questions at issue.

Second, Pietroski seems to shy away from relying explicitly on a distinction between competence and performance. But, certainly, some such distinction must be working in the background. This could easily be applied to explain how CM-semantics can appeal to potentially infinite operations like juncture without ascribing super-human cognitive powers to humans. Semantic competence, as far as composable meanings go, is properly described in potentially infinite terms, if the base is finitely stateable. If this strategy is available to the minimalist, however, it must also suggest itself to the maximalist. This would have

stopped the objection from too much notational power in its tracks. Maximalists are free to argue that the only types ever required in actual linguistic performance will remain below Level Four. Furthermore, large swaths of the types represented at the Levels above Zero will never be used at all. Level One, for example, would have it that there are expressions of type <*e*, *e*>, which are not needed for the semantics of Slangs.

When both theories—minimalism and maximalism—have access to a distinction between competence and performance, direct comparisons of degrees of complexity become more cumbersome. Possibly, still, Pietroski's formulation of different Levels in the Hierarchy of Types could remain useful. But the comparison would have to be between different analyses of one and the same sentence, spelling out the varying operational power required by each analysis. However, if function application and classification are not really different operations, which is still undecided I think, it is not clear which theory wins out on simple sentences. 'Joe snores' is an instruction to fetch two monadic concepts and to M-join them. The resulting monadic concept classifies things into the set of Joe-snoring elements and the set of non-Joe-snoring elements (objects, worlds, or situations). For the maximalist, perhaps, it is an instruction to mentally fetch Joe himself and apply the snoring-function to him, again dividing things into the Joe-snoring elements and everything else. I must admit, in light of the number of different theoretical commitments still open to both theorists, that I have a hard time judging which theory involves more powerful mental operations. All I can say is that focusing on individual sentence or clause types may make the project tractable.

5. *The language of the theorist*

To recap, it seems more difficult to eliminate function application than Pietroski assumes, for the notion is, at least potentially, equivalent to his own notion of classification. We would need a story on which function application and classification must have different realizations in mental mechanisms. Moreover, compositional semantics will always call for notation which is in principle boundlessly powerful, and a competence-performance distinction to explain why only a small part of those representational resources are needed to describe the facts of human performance. Even CM-semantics traffics in notation which allows for boundless productivity but, still, the theory is not exclusively concerned with ideal minds.

It is reasonable to conclude that the most powerful notation on offer is the best one to go by, as long as we can find no hypotheses about cognitive implementations hidden in the notation itself. If so, the Fregean Hierarchy of Types and the lambda-calculus are very good bets. Even if these are powerful, they can be used to describe simple, dumb

operations, like applying *Fx* to some object *a*, in such a manner that the proposition entertained is true if and only if *Fa*.

As I have tried to argue elsewhere (see my 2016, 2019, and *forthcoming*), we should make a distinction between the representational system employed by the theorist to describe and explain the cognition and actions of particular human agents, and the representational system employed by those agents themselves. Human thinkers are often confused in ways that make it inadvisable to simply incorporate their own representations into our explanations. Of course, these representations can always be mentioned, quoted, or otherwise referred to, but they cannot be used directly by the theorist. A simple example is a thinker who confuses the identical twins Bill and Biff and has only a single concept, labeled 'John,' to represent both. Strictly speaking, the thinker then lacks the representational resources to think explicitly about Bill without thinking about Biff at the same time. But still, I have argued, we must be able to ascribe the false belief that Bill is identical to Biff to this agent. This must then be some form of implicit belief.

An opposing view would be that, as theorists, we must be able to reproduce or mirror the confused mental state in our theoretical vocabulary. Well, yes, we must be able to refer to those mental states somehow, but we must also be able to ascribe beliefs which the agent is constitutively unable to represent explicitly. Stuart Hampshire (1975: 123) articulated these two options clearly: 'Perhaps the confusion in his mind cannot be conveyed by any simple account of what he believes: perhaps only a reproduction of the complexity and confusion will be accurate.' So, either our descriptions are reproductions of the blooming, buzzing confusion of our inner lives or they are, rather, models. But if we must sometimes posit the belief that *p* for purposes of explanation even when the explicit thought that *p* is unavailable, we have in effect given up on reproductions.

CM-semantics is in business of providing a model of human minds, just as much as the maximalist alternative. Maximalists can coherently accept any performance-restriction that the minimalist cares to propose. For example, they could agree that human minds can only compose two unary relations or one unary and one binary relation. Nothing in the notation itself disallows the restriction. The real nub of the argument is whether we can find deep, metaphysical differences between conjunctive or predicative composition and function application. Even if I lean one way on this issue right now, I genuinely think it is an open and interesting question.

6. *Conclusion*

One of my underlying themes has been to suggest that, sometimes at least, objections against a compositional semantics for Slangs should also be objections to the same compositional semantics for Mentalese. So, if the objection would have absurd or unpalatable consequences for

Mentalese, perhaps the objection is not reliable in general. Thus, if the objection can be resisted for one of these it can also be resisted for the other. It bears emphasizing, however, that this point certainly depends on various assumptions, one of which is the very notion that we think in a compositional Mentalese.

At one point, Pietroski argues that the Liar paradox is a problem for truth conditional semantics (Chapter 4). Roughly, if my favorite sentence can be 'My favorite sentence is not true,' standard truth conditional semantics will involve contradictions. Pietroski wants to infer that truth conditional semantics should go. If this is correct, it would suffice to prove that Mentalese cannot have a truth conditional semantics either, because my favorite thought can be that my favorite thought is not true. But it just seems too incredible to believe that our thoughts cannot be true or false, and thus have truth conditions. Perhaps we should conclude, then, that there is something wrong with thoughts or sentences of this kind, not that neither thoughts nor sentences can be true or false.

The broader point is to suggest that the expressive power of our theoretical vocabulary is not, as such, a reliable indicator of the explanatory power, or lack thereof, of any theory expressed by that vocabulary. Working formal semanticists tend to avoid making proclamations about where human linguistic competence ends and god-like mental powers would have to start. But surely, Pietroski is correct to point out that Level Four in the Fregean Hierarchy of Types is not needed to explain competence in Slangs. My argument is, basically, that one need not be a CM-semanticist, that is, one need not eliminate function application, to make this particular point.

References

Dowty, D. R., Wall, R. E. and Peters, S. 1981. *Introduction to Montague Semantics*. Dodrecht: Springer.

Frege, G. 1956. "The thought: A logical inquiry." *Mind* 65 (259): 289–311.

Glanzberg, M. 2014. "Explanation and Partiality in Semantic Theory." In A. Burgess and B. Sherman (eds.). *Metasemantics*. Oxford: Oxford University Press: 259–292.

Hampshire, S. 1975. *Freedom of the Individual*. 2nd ed. London: Chatto & Windus.

Heim, I. and Kratzer, A. 1998. *Semantics in Generative Grammar*. Oxford: Blackwell.

Lewis, D. K. 1975. "Languages and language." In K. Gunderson (ed.). *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press: 3–35.

Pietroski, P. 2018. *Conjoining Meanings: Semantics without Truth Values*. Oxford: Oxford University Press.

Roberts, C. 2018. "Speech Acts in Discourse Context." In D. Fogal, D. W. Harris, and M. Moss (eds.). *New Work on Speech Acts*. Oxford: Oxford University Press: 317–359.

Shieh, S. 2019. *Necessity Lost: Modality and Logic in Early Analytic Philosophy. Volume 1*. Oxford: Oxford University Press.

Unnsteinsson, E. 2016. "Confusion is Corruptive Belief in False Identity." *Canadian Journal of Philosophy* 46 (2): 204–227.

Unnsteinsson, E. 2019. "The edenic theory of reference." *Inquiry: An Interdisciplinary Journal of Philosophy* 62 (3): 276–308.

Unnsteinsson, E. Forthcoming. *Talking About*.

Croatian Journal of Philosophy
Vol. XX, No. 60, 2020

Generative Linguistics Meets Normative Inferentialism: Part 1

DAVID PEREPLYOTCHIK*
Kent State University, Ohio, USA

This is the first installment of a two-part essay. Limitations of space prevented the publication of the full essay in present issue of the Journal. The second installment will appear in the next issue, 2021 (1). My overall goal is to outline a strategy for integrating generative linguistics with a broadly pragmatist approach to meaning and communication. Two immensely useful guides in this venture are Robert Brandom and Paul Pietroski. Squarely in the Chomskyan tradition, Pietroski's recent book, Conjoining Meanings, offers an approach to natural-language semantics that rejects foundational assumptions widely held amongst philosophers and linguists. In particular, he argues against extensionalism—the view that meanings are (or determine) truth and satisfaction conditions. Having arrived at the same conclusion by way of Brandom's deflationist account of truth and reference, I'll argue that both theorists have important contributions to make to a broader anti-extensionalist approach to language. What appears here as Part 1 of the essay is largely exegetical, laying out what I see as the core aspects of Brandom's normative inferentialism (§1) and Pietroski's naturalistic semantics (§2). In Part 2 (next issue), I argue that there are many convergences between these two theoretical frameworks and, contrary to first appearances, very few points of substantive disagreement between them. If the integration strategy that I propose is correct, then what appear to be sharply contrasting commitments are better seen as interrelated verbal differences that come down to different—but complementary—explanatory goals. The residual disputes are, however, stubborn. I end by discussing how to square Pietroski's commitment to predicativism with Brandom's argument that a predicativist language is in principle incapable of expressing ordinary conditionals.

Keywords: Generative linguistics, anti-extensionalism, normativity, inferentialism, predicativism, public language, communication.

* My deepest thanks to Daniel Harris, Ryan DeChant, Elmar Unnsteinsson, Eliot Michaelson, and Jessica Lynne Cooperrider for helpful advice on earlier drafts

Introduction

I take it that the correct approach to natural-language syntax is the one that Noam Chomsky outlined as early as the 1950s and, along with many others, has continually refined over the past seven decades. The ongoing research program of generative linguistics that his syntactic theorizing inspired has, in the fullness of time, yielded a diversity of impressive results. These include exciting and previously unimaginable empirical discoveries about the human capacity for language, both in broad scope—e.g., recursive generability and the principles-and-parameters model—and at the level of fine-structure (e.g., traces, parasitic gaps, etc.). However, as a theorist interested not only in syntax but also in semantics, I find myself in a difficult and somewhat awkward position.

Not to complain, but, you see, I happen to have learned my semantics from the work of Robert Brandom, and it's safe to say that I've drunk the Kool-Aid that he served up in his magnum opus, *Making It Explicit*. Having thus bought into *both* Chomsky's generative grammar *and* Brandom's normative inferentialism, I now find myself facing the daunting challenge of bridging the apparent chasm between the two. It may be that I'm utterly alone in this quandary, but in the course of the present discussion, I hope—perhaps somewhat perversely—to draw others into it as well.

There may never be an academic conference addressing the common themes and shared commitments of generative linguistics and normative inferentialism. For all that, the differences between them are, I believe, more boring—i.e., verbal or sociological—than is widely assumed. In what follows, I'll argue that, contrary to first appearances, there are actually very *few* points of substantive disagreement between them. The residual differences are stubborn, to be sure, but this can only be appreciated after a suitably wide collection of background agreements is put into place. I devote the first half of the present discussion to this latter task.

Squarely in the Chomskyan tradition, Paul Pietroski's recent book, *Conjoining Meanings*, offers an approach to natural-language semantics that rejects foundational assumptions widely held amongst philosophers and linguists. In particular, he argues against the view that

of this paper; thanks also to Nate Charlow, Matt Moss, and other participants of the 2018 Ontario Meaning Workshop, for stimulating exchanges on related topics. That said, if some of what I claim here is false, then these people are all partly to blame. Ditto for cases where I say something awkwarder than usual; they should of caught that. My sincere gratitude, lastly, to all of the participants at the two symposia at which I was fortunate to present earlier versions of these ideas, including the 2018 Croatian Meaning Workshop. From the latter group, I owe special thanks to Paul Pietroski (obviously), Michael Devitt, John Collins, Anna Drożdżowicz, and, most of all, to Dunja Jutronić—not only for inviting me to participate in the workshop, but for her saintly and near-inexhaustible patience with me throughout this project. Many thanks also to Nenad Miščević for allowing the second half of the present essay to appear in the next issue of the Journal.

meanings are, or even determine, extensions. The latter include such familiar semantic properties as truth conditions, satisfaction conditions, and denotation/reference. Having arrived at the same conclusion by way of Brandom's deflationist account of truth (§1.3), I began to glimpse the possibility of a fruitful merger of ideas. The present essay represents a first pass at integrating the generative linguist's empirical insights about human psychology with the broadly pragmatist framework about mind and language. I'll argue that both have important contributions to make to our overall understanding of language. The easy part is spelling out what; the harder part is assessing the residual disagreements.

Here's the overall plan:

In §1, I survey a range of core commitments that jointly constitute Brandom's philosophical project—most centrally, his normative pragmatics, inferentialist semantics, and substitutional syntax. Along the way, I note his intellectual debts to David Lewis, including the large-scale explanatory goals that animate Brandom's inquiry. These, I later argue, are in many ways orthogonal to Pietroski's concerns. The latter claim figures in my broader argument that the two approaches can be fruitfully combined, potential protests from both sides notwithstanding.

In §2, I outline Pietroski's position, focusing on his explanatory aims, his empirical methodology, and the substance of his proposal for a theory of human semantic competence. I take up Pietroski's arguments against Lewis's approach to natural language, with the aim of showing that Brandom's theoretical goals differ sufficiently from Lewis's to inoculate him against Pietroski's criticisms. Turning to Pietroski's discussions of Frege, I point out that his work in cognitive science undercuts some of Brandom's claims in "Why Philosophy Has Failed Cognitive Science". As we'll see, far from ignoring Frege, Pietroski incorporates many of his insights into an empirical account of psychological processes. Nevertheless, some of the Fregean lessons that Brandom emphasizes do *not* seem to have moved Pietroski. In later sections, I explore some possible reasons for this.

This will conclude Part 1 of the essay. Limitations of space prevented the publication of the full essay in present issue of the Journal. The second installment will appear in the next issue, 2021 (1). Here is a preview of what it contains:

I devote §3 to a survey of the core commitments that Pietroski and Brandom have in common. As already noted, both reject truth-conditional semantics and seek to develop alternative frameworks for semantic theorizing. Similarly, I'll point out, the alternatives they propose can both be seen as taking referential *purport* to be the distinctive feature of language. This contrasts with received views in (meta) semantics that focus, in the first instance, on referential *success*, leaving reference *failure* (e.g., empty names) for special treatment, as a blessedly rare "defective" case. This is related, I suspect, to the further convergence between the Brandom and the Pietroski on the proper

treatment of the distinction between *de dicto* and *de re* constructions. Last, but no less important, is their common rejection of the idea that communication requires an *identity* between the meanings expressed by a speaker and those grasped by the hearer. I'll note that this shared commitment undercuts some of the main arguments against meaning holism—an inherent feature of Brandom's inferentialist account.

A discussion of the differences between Pietroski and Brandom occupies the remainder of the essay. At first glance, the contrast between them seems as vivid as any in the field. Brandom's project is explicitly *normative*, pursued largely from the armchair, and aims to provide an account of concept users *as such*—not just humans, but *any* conceivable concept-mongering creature (or artificial system). By contrast, Pietroski's project is avowedly *descriptive*, constrained by *empirical* data, and aims to provide an account of *actual humans*—particularly, the "Slangs" that they naturally produce and consume. (Pietroski uses 'Slang' as a catch-all term for the natural languages that children acquire.) These differences ramify quickly. For instance, Brandom's focus on social practices of communication seems to be at odds with Pietroski's "individualistic" methodology, which is characteristic of generativist views more broadly (Chomsky 2000; Collins 2012). This contrast seems particularly sharp in light of Brandom's commitment to the existence of (something like) public languages—at least in the sense of productive and flexible norms governing the communal practice of "giving and asking for reasons" (GOGAR). All of these issues are discussed in §4, with the aim of gradually blunting the force of what initially appear to be quite sharp differences.

If one insists on seeing Brandom's and Pietroski's inquiries as targeting a common subject matter, then one is sure to view the differences between them as substantive theoretical *disagreements*. They do, after all, use the same term, 'language', which at least *suggests* that they're talking about one and the same phenomenon. But this way of viewing the situation is optional, at best. We've learned by now to stop assuming that theorists are targeting the same phenomenon simply on account of their using homophonous terms. Two theorists can press the same bit of folk vocabulary, e.g., 'meaning' or 'concept', into more weighty theoretical labor in quite different ways. That being so, one can just as well see the inferentialist and the generativist as addressing *different* (though undoubtedly related) topics—each providing insights about his chosen domain of inquiry, and leaving the rest of us to wonder how those insights might be integrated, or at least brought to bear on one another. This is the strategy I'll recommend throughout the present discussion.

In arguing that the two are, at the very least, logically *compatible*, one owes an account—or, at a minimum, a broad *sketch* of an account—of the theoretical *relation* between them. The proposal that I'll develop is that Brandom's explanatory ambitions differ from Pietroski's in precisely the ways that are paradigmatic of "*inter-level* theoretical

relations". My suggestion is that Brandom is attempting to furnish a *high-level description* of a quite general phenomenon—*language use, as such*. This presupposes that lower-level implementations of the relevant generalizations can vary widely.[1] Pietroski's theoretical aims are, though certainly more exciting from an empirical standpoint, a good bit narrower than Brandom's, in that they deal exclusively with the human case. As with any lower-level account of a more general phenomenon, Pietroski's view is compatible, in principle with any number of higher-level descriptions of language. Thus, while it's incorrect—or, at any rate, misleading—to say that the two accounts are strictly orthogonal to one another, the fact is that each places very few constraints on the other. It *should* be surprising, then, when we find substantive points of contact between them, whether these be points of contention or convergence.

Still, even if my suggestion is right that the two frameworks are more compatible than might initially appear, we must face up to the residual differences that credibly threaten my reconciling project. The most stubborn of these, which I accordingly leave to the very end, has to do with Pietroski's arguments for a specific version of *predicativism*—roughly, the view that all subsentential entities are best treated as predicates, not (say) singular terms (§2.4). To be sure, Pietroski's commitment to this view is not a central aspect of the overall generativist enterprise. Rather, it's a tendentious empirical hypothesis, for which he offers correspondingly forceful arguments. That being so, if it were to turn out that the hypothesis is false, generative linguistics would go on without a hitch. Still, I focus on this issue because it raises much larger questions about how to treat subsentential entities, not just at the level of semantics, but also at the level of *syntax*.

The notion of syntax that Brandom employs is *substitutional*—not in the sense of "substitutional quantification" (though he endorses that too, on independent grounds), but, rather, in the sense that he takes *sentences,* i.e., expressions of *full thoughts*, to be the primary vehicles of meaning. On his view, subsentential items (words, morphemes, etc.) are the products of taking a "substitutional scalpel" to the antecedently-interpreted *sentential* unities. We'll take a first stab at unpacking the scalpel metaphor in §1.5, and then come back to it in more detail in §4.3. While this substitutional approach *may* have a home in a Fregean semantics—Pietroski's powerful arguments notwithstanding—there is no *obvious* sense in which it can be legitimately applied to the syntax of human languages.

The trick that I pull repeatedly throughout the second half of this essay—namely, that of relegating the two inquiries to distinct theoretical "levels"—doesn't get much of a grip here. For, it's a requirement of any such picture that an account pitched at the (relatively) higher level

[1] Indeed, as Fodor (1975) points out, without substantive constraints from the higher-level account, such "realization bases" might differ *indefinitely*.

of analysis should be *compatible*, at least in *core* respects, with any lower-level account of the "realization base". But Brandom's substitutional syntax seems, even upon close scrutiny, to be not only *different* from generative grammar, but decidedly *at odds* with it. I strongly suspect that this different comes down to a methodological conflict between the generativist's "bottom-up" treatment of subsentential items and the inferentialist's "top-down" alternative.[2] In §4, I discuss and evaluate some ways of viewing this disagreement, arriving ultimately at the following bittersweet conclusions.

Despite strong grounds for optimism about the possibility of integrating normative inferentialism with the up-and-running research program of generative linguistics, it must be admitted that rendering Brandom's substitutional approach to syntax compatible with the going theories in generative grammar (e.g., Chomsky's minimalist program) presents an obstinate challenge. Though I can think of no reason that the challenge is insuperable *in principle*, it's nevertheless the case that I am not, at present, equipped to meet it myself. Perhaps others can do better in this regard—a task I invite, encourage, and exhort philosophers of language to undertake, on the strength of the positive arguments adduced here.

1. *Brandom: From normative pragmatics to inferentialist semantics (and back)*

Introduction

Robert Brandom's philosophical project is grand in both scope and ambition.[3] The resulting theoretical framework has a number of moving parts, to put it mildly. In this section, I'll lay out what I take to be the central commitments of his "normative inferentialism", with particular focus on those that pertain to the broader goals of this discussion—i.e., the proposal to integrate the generativist and inferentialist research programs.

Although this section is intended to be largely exegetical, postponing critical evaluation to §4, I should emphasize that it wouldn't matter to me very much if Brandom wouldn't put things *quite* the way that I do below. While his account is by far the most well-worked out version of normative inferentialism, and hence immensely useful as a guide

[2] See Collins (2012) for a detailed and rewarding discussion of this issue. The contrast that Collins draws between sentence-first and word-first approaches (my terms, not his) serve, if anything, to *sharpen* the contrast that I'm worried about here. My hunch is that coming to grips with Collins' conclusions in that work will be crucial to resolving the issues I raise in §4.3.

[3] The picture I sketch in what follows is drawn mostly from the material in Brandom (1994, 2000, and 2008). In what follows, I'll occasionally abbreviate these to *MIE* (*Making it Explicit*), *AR (Articulating Reasons)*, and *BSD (Between Saying and Doing* 2008*)*, respectively. The final sections of this essay also make heavy use of the material in "Why Philosophy Has Failed Cognitive Science" (in Brandom 2009).

in this area of inquiry, my intent is not so much to capture every nook and cranny of one particular theorist's gargantuan philosophical system. Rather, the goal of this section is to present and motivate an account of language that is attractive enough to warrant comparison with Pietroski's independently attractive proposals about natural-language semantics (§2). Ultimately, it's only by comparing them that we can put ourselves in a position to clearly assess the merits of either, let alone to contemplate their integration.

1.1. *Methodology and explanatory aims*

The central questions, for Brandom, are about what *constitutes* language use—not by humans, necessarily, but by any natural creature or artificial system to which attributions of a linguistic competence are warranted. Given that normal adult humans have mastery of at least one natural language, Brandom's account will, in *some* sense, apply to us as well—though perhaps only as a particular instance of a much more general phenomenon. Still, the inquiry he undertakes is not a straightforwardly psychological one; nor is it pitched as a historical or anthropological hypothesis. Rather, the idea is that, armed with a philosophically sophisticated and conceptually articulated account of "what the trick *is*"—where "the trick" is language use, broadly construed—an empirical scientist (a linguist, psychologist, or artificial intelligence researcher) can ask more detailed questions about "how the trick *happens to be done*" by one or another creature or artifact.

This latter kind of research is bound to yield an increasingly refined picture of some particular *type* of linguistic competence—paradigmatically, the human type (though one often hears of impressive progress in interpreting the languages of other social creatures, e.g., prairie dogs and dolphins). Thus, although the account is intended to apply far beyond the human case—to aliens, robots, and other terrestrial animals—there is no commitment on Brandom's part to the effect that empirical findings can have no bearing *whatsoever* on his philosophical claims. Nor does he hold that legitimate empirical inquiry can take place only *after* a credible philosophical account has been supplied. This is just one sense in which normative inferentialism and empirical science, including generative linguistics, are not in competition.

One might wonder, at this point, how far removed Brandom's project is from empirical considerations. Perhaps *too* far? True, many of the thinkers with whom his work most directly engages did *not* conceive of their inquiries on the model of empirical theorizing. In fact, the three philosophers who figure most centrally in *MIE*—namely, Kant, Frege, and Wittgenstein—all famously made a point of distancing themselves from natural science. But we have to keep in mind, here as elsewhere, that Brandom's theoretical framework is "a house with many rooms". Thus, we need not take a one-size-fits-all approach to this question; indeed, there are good reasons not to.

In providing detailed analyses of linguistic constructions in distinctively human languages—*de dicto/re* reports, truth-talk, (in)definite descriptions, indexicals, deixis, anaphora, modals, quantification, predicates, and singular terms—Brandom relies on exactly the same stock of empirical considerations that one finds in standard semantics texts (e.g., Heim and Kratzer 1998). Such data are often cast in terms of intuitive judgments concerning the *truth*-values or *truth*-conditions of a target sentence, often in the context of auxiliary assumptions that are supplied by an accompanying vignette. But these very same data can equally well be recast as competent speakers' judgements concerning (not truth but) *inferential proprieties*—e.g., what one is required or permitted to infer on the basis of the assumptions supplied (in combination with all prior background information that is assumed to be common-knowledge).[4]

A similar treatment can be given of metalinguistic intutions/judgments concerning minimal pairs and sentential ambiguity. Two sentences differ in meaning, on the inferentialist account, just in case they differ in respect of the inferential proprieties that govern their use. Likewise, a sentence is ambiguous just in case it is capable of playing two distinct/incompatible inferential roles. And, though he doesn't, to my knowledge, ever discuss the phenomenon of polysemy, I presume that Brandom would treat is as a case of *overlapping* inferential properties—perhaps ones that meet some further normative or inferential conditions.

The fact that Brandom often appeals to precisely such data in developing his inferentialist semantics suggests that at least *this* aspect of his view is firmly grounded in empirical fact. But, as I'll emphasize later, Brandom's use of these data is not empirical but, rather, *illustrative*. He is, in other words, using examples from English—indeed, almost exclusively English, in contrast with the generativist's cross-linguistic methodology—as case studies that, according to him, exemplify more general phenomena. Thus, the English-language examples that he occasionally provides—a bit too rarely, one might lament—serve not as empirical data for a scientific or naturalistic hypothesis. Rather, the most charitable reading of his appeals to such examples casts them as attempts to help us to get a conceptual grip on features of language(s) *as such*.

Still, it must be admitted that other aspects of Brandom's overall picture are far less tethered to the facts on the ground. Presently, we'll see that his normative pragmatics—to which his downstream proposals, including inferentialism, are conceptually subordinate—makes only very minimal empirical assumptions. For instance, it takes for granted that complex social creatures came into being *somehow or other*—e.g., via evolution by natural selection, or by deliberate engi-

[4] My thanks to Eliot Michaelson and Daniel Harris for helping me to see the points in this paragraph more clearly.

neering, as with AI. Brandom likewise presumes that the behavioral control systems of such creatures/artifacts—brains, motherboards, or what have you—work *somehow or other*; else they couldn't perform the behaviors that institute social norms, given reasonable naturalistic constraints. Still, putting aside such near-vacuous assumptions, this aspect of Brandom's project can fairly be described as an armchair enterprise.[5] Whether that's something to hold against it—e.g., on the basis of one or another naturalist scruple—is something we can only ascertain only after we've surveyed the details of his overall proposal. To these we now turn.

1.2. *Normative pragmatics*

Brandom begins by situating all linguistic practices in the wider realm of activities that are, in some sense, *rule-governed*. The notion of a rule plainly stands in need of careful articulation. Given our philosophical history, two options immediately suggest themselves—what Brandom calls "regulism" and "regularism". According to the *regulist*, rule-following is a matter of obeying rules that one can *explicitly* formulate or comprehend. By contrast, the *regularist* holds that rule-following is a matter of being disposed to behave in a way that accords with one or another empirical *regularity*. Brandom rejects both of these alternatives, though, as we'll see, his own account is an attempt to split the difference between them.

Regulism is vitiated, he argues, by the fact that obeying an explicit rule—e.g., a dictionary definition or an academic/prescriptive convention of grammar or style—requires first *interpreting* the rule. This, in turn, requires *deploying concepts*—a version of precisely the phenomenon that we're seeking to explain. The regularist option, he contends, faces a distinct challenge. A pattern of behavior—whether finite or infinite, actual or potential—can be either a *successful* or a *defective* case of following a given rule. So, with regard to *any* pattern of performance, the question always stands: has the rule in question been followed *correctly*? Assessments of correctness are, on Brandom's view, inherently normative. Yet, what the regularist offers is a purely descriptive account, couched in the language of cognitive or behavioral dispositions (and perhaps other, related alethic modal notions; see §3.1).

Brandom's positive proposal insists on the normativity of assessment, characteristic of the regulist view, but jettisons the requirement that the rules in question be manifested explicitly from the very outset. Though rules of practice can *eventually* come to be articulated—i.e., "made explicit"—by a community of concept-using creatures, such rules are, in the first instance, *implicit* in the social practices of the creatures in question.

[5] Notably, a former student of Brandom's, John MacFarlane, has programmed a version of "the game of giving and asking for reasons" (GOGAR) for the popular online game, *The Sims*. https://www.johnmacfarlane.net/gogar.html

Why social? Why not, instead, the practices of an individual? Put crudely, the reason is that isolated individuals cannot, in principle, serve as a normative check on their own judgment and behavior. In order for a creature to be so much as *subject* to normative assessment, its behavior must take place in a context where other creatures respond to it in ways that signal social (dis)approval. How numerous and long-lasting must the social relations be in order to institute a social practice, properly so called? Brandom's answer, which will become relevant later in the discussion (§4.1), may be somewhat surprising. Terms like 'social' and 'communal' bring to mind a relatively large group of creatures. But, when he presses these folk terms into theoretical service, Brandom's official view is far less committal than that. All it really requires is a *dyadic* "I-thou" relation—i.e., a case of mutual recognition, in respect of authority and responsibility, on the part of at least *two* creatures/systems. Such relations of mutual recognition can be merely *implicit* in the 2+ creatures' overt practices toward one another—e.g., one or another type of *social sanction*.

In the most basic case, social sanctions come down to either naked violence or the provision of necessities—beatings and feedings. But once a social practice becomes sufficiently complex, it comes to include not only such "external" sanctions, but also "internal" ones—e.g., initially, the granting of privileges and later the exchange of *tokens* of privilege. (We are invited here to imagine a special fruit that permits its bearer to enter a particular territory without being attacked by the creatures who guard it.) With each additional layer of interwoven internal sanctions, the community becomes increasingly ripe for instituting not merely norms of practical action, broadly construed, but the more specifically *linguistic* norms of assertion and demand. I include the latter here, not because Brandom ever treats it in detail, but on account of his frequent invocation of the trope "the game of giving *and asking for* reasons" (my emphasis, to be sure). While the "asking" part seems to deserve equal attention, Brandom takes the norms governing *assertion* to be the "downtown" of language—a backhanded adaptation of Wittgenstein's metaphor of language as a city with *no* downtown.

Having argued that assertion is the fundamental pragmatic notion, Brandom goes on to give an account of it in terms of the normative social statuses of commitment and entitlement. To illustrate these notions, let's work through a hypothetical example of how the game of giving and asking for reasons might be played amongst a group of primitive hominids—or, for that matter, current prairie dogs.[6]

Suppose a creature produces a public token in a social context, and that this act has—if only in that context—the pragmatic significance of explicitly committing the creature, in the eyes of its community, to its

[6] I am acutely aware that, despite my efforts at rendering the scenery in a plausibly naturalistic light, the example is not only fictional but transparently artificial in countless respects. I trust these won't matter for the sake of making the key points accessible.

being the case that the enemy is approaching. Each of the other creatures in the group evaluates this commitment, assessing it as correct or incorrect on the basis of their own commitments, explicit or otherwise. In doing so, they take a *normative stance* toward the "speaker", whom the group might then treat as being *entitled* to that commitment. The speaker can be entitled to a claim either by default—e.g., in cases of joint perception or contextually-relevant common knowledge—or by having undertaken prior commitments that jointly warrant the one in question. Initially, such normative attitudes are implicit in the assessors' overt treatment of the speaker. The group might, for instance, shift its attention in the direction of the speaker's gaze upon hearing 'Enemy!". If the enemy is indeed approaching in a way that is perceptually evident to the group, the entitlement is thereby secured.

Suppose now that the same speaker, now quite anxious, produces another token—e.g., "Run!" This counts not only as an explicit commitment to a (potentially joint) plan of action, but also an implicit commitment to the goodness of the inference from "Enemy!" to "Run!" And while the members of the group assessed the speaker as being entitled to the first claim, they may well go on to treat this new explicit commitment, i.e., "Run!", as patently unwarranted in the circumstances. Again, in the most basic cases, this "treatment" or "stance" toward the speaker can take the form of overt actions on the part of other group members—e.g., grabbing hold of the speaker and keeping them in place. Another form of response might be the production of overt tokens that commit the group, including the original now-frightened speaker, to an incompatible plan—e.g., "Stay!" and "Fight!". This normative incompatibility is itself implicit in the overall communal practices of the group.

From these primitive beginnings, Brandom suggests, a practice can evolve in such a way as to allow for speakers to make *explicit* their commitments regarding the goodness—or, in his terms, "material propriety"—of the inferences that were previously only *implicit* in their practices. Paradigmatically, this is achieved by introducing an expression that has the significance of a conditional. For instance, we can imagine a newly-evolved creature—call it v2.0—that has achieved what Brandom calls "semantic self-consciousness". This involves discursively representing not just enemies and escape strategies, but also *inferential relations between claims*. Creature v2.0 can make explicit its commitment to the goodness of a particular inference by producing a token that has the significance of the conditional, e.g., "If Enemy then Run!"[7] Further elaborations of v2.0's language are manifested with the introduction of new bits of logical vocabulary, all of which serve to express commitments regarding various inferential relations. For instance, negation expresses the relation of inferential *incompatibility* between claims (see below). This is the core thesis of what Brandom calls "logical expressivism".

[7] Note that the force operator, '!', is stripped off from the atomic propositions, "Enemy!" and "Run!", when the latter are embedded in a conditional. This point comes to the foreground in §4.2.

Turning to more complex logico-semantic devices, consider the phenomenon of indirect discourse. Brandom proposes that *de dicto* reports of "what was said" are used to make explicit the *attribution* of commitments to oneself or to others. For instance, my asserting "Dan said that p" makes explicit *my* commitment to Dan's having explicitly undertaken *his own* commitment to the effect that *p*. In a similar fashion, epistemic vocabulary can play the role of making explicit a speaker's *assessment* of someone's *entitlement* to the commitments that they've undertaken. For instance, were I to upgrade my assertion to "Dan *knows* that *p*," I would not only be *attributing* the commitment to Dan, but also committing *myself* to it, and explicitly representing him as being entitled to it. Brandom (2001: ch. 4) points out that these three *pragmatic* aspects of knowledge *attributions*—entitlement-attribution, commitment-undertaking, and commitment-attribution—correspond, in that order, to the three elements of the traditional Justified-True-Belief account of knowledge.

What about deontic modal terms? Brandom argues that these serve to make explicit one's commitment to a plan of action. In saying "I ought to φ", I make explicit my commitment to a plan to φ. Similarly, terms like 'must' can be used to make explicit an inferential propriety that is insensitive to changes in auxiliary assumptions, up to some boundary condition—e.g., natural or legislated laws. Thus, assertions like "In order to light the wick safely, one must first clean one's hands" serve to make explicit a speaker's commitment to the material propriety of the inference from "The wick-lighting is safe" to "Your hands are clean," irrespective of what commitments they have concerning a wide range of possible auxiliary assumptions, such as "It's raining elsewhere," "I never met my grandfather," and "Child-trafficking is a serious problem." Whether or not these latter are included in one's set of commitments, the inference from "The wick-lighting is safe" to "Your hands are clean" is ostensibly good. Of course, auxiliary commitments that conflict with one's views on *natural laws*—e.g., "Gasoline burns when lit"—would render the inference materially invalid. That's what talk of "boundary conditions" is intended to capture. In the case of the purely nomic reading of 'must', as in "Oxygen must be present for combustion," the inference from "Combustion is occurring" to "Oxygen was present" is good under substitution of virtually any commitment, other than those regarding physical or chemical law.

In this vein, Brandom provides rich pragmatic analyses of a variety of other "vocabularies", including the (meta)semantic devices 'represents' and 'is about', as well as indexicals, alethic modals, anaphoric pronouns, and *de re* attitude reports. Some of the details of these analyses will emerge throughout the present discussion, and I'll devote special attention to his account of *de re* constructions in §3.4. For the moment, the case that's most important to examine is that of 'true', as this bears directly on Brandom's rejection of truth-conditional semantics—arguably the main negative contention that he shares with Pietroski.

Brandom offers a clear and well-motivated alternative to the standard treatment of truth in terms of "correspondence"—a notoriously vexed notion that lay at the heart of truth-conditional semantics. Instead he develops a refined version of the "deflationary" approach of truth and reference—arguably the best on the market, as we'll see presently.

1.3. *Deflationism about truth and reference*

In asserting that things are thus-and-so, a speaker takes on the pragmatic normative status of a discursive commitment. What is it, then, for *another* person to say that the first speaker's assertion is *true*? The normative pragmatic answer is this: in asserting that some claim is true, one not only *ascribes* a commitment to the speaker who made it, but one also *undertakes that commitment oneself*.[8] This allows for the possibility—enormously useful in social practice—for a speaker to take on commitments that they cannot at present articulate. The inability may be either to memory loss ("I don't remember exactly what she said, but it was definitely true") or to time constraints ("She gave a long speech; I'm not going to repeat the whole thing now, but everything she said was obviously true"). In the most interesting cases, complete articulation is impossible within physical limits, the set of commitments being literally infinite, as with "The theorems of Peano arithmetic are all true".

Thus, on Brandom's view, the term 'true' and its cognates ('truth', 'correctness', etc.) all serve a distinctive *expressive* function, without which branches of discourse such as mathematics would be impossible. Specifically, these terms all serve to express a commitment to something already asserted (or, at any rate, assert*ible*). Brandom thus labels this an "expressivist" account of truth, of a piece with his more general expressivist approach to logical vocabulary. The semantics for truth ascriptions is elaborated still further in light of his discussion of anaphora (§1.7). The notion of inter-sentential (and inter-speaker) anaphora will allow us to appreciate how truth ascriptions can have the pragmatic function of allowing the *inheritance* of commitments and entitlements across inter-personal exchanges.

If this pragmatic expressivist account of our use of 'true' (and related terms) is correct, then there is no obvious reason why anything further needs to be said about *truth*. The latter is often conceived of as a metaphysical language-world relation—the very one denoted/satisfied by the term 'truth' and its cognates. But there seems to be no explanatory work for which such a relation is *obviously* indispensable—neither in semantics nor, Brandom argues, in any other area of theorizing. This puts his view in the same camp as other versions of "deflationism" about truth—particularly, the well-known disquotational (Quine

[8] In the special case where the speaker and the assessor are identical—as in, "What I'm saying is *true*!"—the commitment is both redundant and guaranteed, though the term is still useful, in such cases, if only for emphasis and the like.

1970), minimalist (Horwich 1990), and prosentential accounts (Grover, Camp, and Belnap 1975). I take all of these to share the following core commitments:

(i) a rejection of any account/analysis of truth in terms of correspondence, coherence, or warranted assertibility, on the grounds that truth is *not a relation* (of *any* kind)

(ii) the aim of casting the notion of meaning as *explanatorily prior* to that of truth, both in semantics and elsewhere (including metaphysics, epistemology, and ethics)

The differences between disquotationalism, minimalism, and prosententialism have mostly to do with matters of detail, such as whether to ascribe truth to sentences or to propositions, or how exactly to interpret Tarski biconditionals, liar sentences, and quantified truth ascriptions. These disputes are all strictly irrelevant for our purposes. What's important here is that Brandom's version of deflationism is designed to claim the virtues of each of these prior accounts, without succumbing to the technical objections that have been lodged against them. The three main improvements he suggests are (i) subordinating the semantics of truth ascriptions to his brand of normative pragmatics, (ii) paying closer attention to the *syntax* of truth ascriptions, especially their intersentential anaphoric structure (§1.7), and (iii) extending the deflationist account to other semantic notions, including reference, satisfaction, and *de re* representation. While (i) is a straightforward application of Brandom's broader strategy, and (ii) serves largely to immunize his version of deflationism from extant objections, (iii) strikes me as a genuine extension of Brandom's normative pragmatics, allowing it to handle both sentential *and sub*sentential expressions.

The notions of truth and reference are plainly central to the project of truth-conditional semantics. Thus, many have noted that a deflationist account of these notions requires a radical re-thinking of what shape a formal semantic theory should take. In this regard, we now have an embarrassment of riches. In addition to old-school proposals about warranted assertibility, and the pragmatists' short-lived "success semantics" (see Brandom 2009 for critique), we now have the benefit of more modern proposals, including both Pietroski's cognitivist account (§2) and Brandom's inferentialism. Let's examine the latter.

1.4. *Inferentialist semantics*

Having situated assertional practices within the broader sphere of rule-governed social activity, Brandom has introduced his key *pragmatic* notions of commitment and entitlement. He goes on to show how these normative statuses, taken together, can be used to construct a *semantic* theory, whose business it is to explain (in some sense) *how* linguistic expressions can come to play the roles that they do in a community's assertional practices.

In familiar fashion, the explanation goes by way of assigning "meanings" or "semantic values" to expression types. But, in keeping with his other commitments, Brandom does *not* equate meanings with truth conditions, sets of possible worlds, pragmatic success conditions, or assertibility criteria. Rather, he subordinates his semantic theory to the normative pragmatics just outlined, by treating meanings as the *inferential proprieties* that govern the use of linguistic expressions. Slurring over a considerable mass of detail, we can summarize the proposal as follows:

Inferentialism: For a given propositional expression, 'P', the meaning of 'P' can be modeled as the set of sets of other propositional expressions that

(i) entitle one to 'P' in the presence of (various sets of) auxiliary commitments,

(ii) commit one to 'P' in the presence of (various sets of) auxiliary commitments, as well as those to which

(iii) 'P' commits one in the presence of (various sets of) auxiliary commitments, and

(iv) 'P' entitles one in the presence of (various sets of) auxiliary commitments.

A particularly useful *compound* inferential relation turns out to be that of *incompatibility*, wherein taking on one commitment precludes a speaker from becoming entitled to another. In this sense, a commitment to "Herbie is a dog" is incompatible with entitlement to "Herbie is a bird". Brandom (2008: ch. 5) shows how to build a *modal* propositional semantics on the basis of just this incompatibility relation, treating the negation of a claim, for instance, as the minimal set of commitments that are incompatible with it. Here again, the details are illuminating, but only one significant upshot bears highlighting for present purposes.

Casting the meaning of an expression in terms of its inferential proprieties vis-à-vis *other* expressions plainly commits one to meaning holism. A common charge against theories of a holist stripe is that they founder on the rock of compositionality. For instance, Fodor and Lepore (1992) famously argue that inferential roles don't compose, whereas meanings do; *a fortiori*, meanings can't be inferential roles. But the formal incompatibility semantics developed by Brandom (2008) provides a direct counterexample the main premise of this argument, by demonstrating how a inferentialist semantics *can* in fact *provably* meet reasonable compositionality constraints, at least in the modal propositional case.[9] In any event, we will see that there are other reasons to reject Fodor and Lepore's arguments.

[9] Although, as of this writing, an analogous proof for the *quantificational* case remains elusive, I am aware of no principled reasons for thinking that such a proof won't emerge—if not tomorrow, then *someday*. As will become clear throughout, I adopt a resolutely optimistic attitude toward such matters.

1.5. *Substitutional "syntax"*

We've now put on the table both a normative pragmatics and an inferentialist semantics. However, it's relatively uncontroversial that only *proposition*-sized expressions can enter *directly* into inferential relations as premises and conclusions.[10] That being so, we still need to say how *sub*sentential bits of language can have meanings of their own. Identifying subsentential expressions will allow us to explain how such expressions can go on to contribute to the indefinitely many assertions that a creature like us can interpret and produce. While there is no *conceptual* barrier, on Brandom's picture, to a community of creatures/robots using a language with only *finitely* many complex expressions, our own case plainly illustrates that languages can and do come in varieties that admit of *productive* generation. So, while a first-pass presentation of the inferentialist approach is best conducted in terms of a community of creatures that uses a finite language—such as might easily be found in (extra)terrestrial nature or constructed in a robotics laboratory (e.g., AIBO dogs)—it does *not* follow, and is not true, that the inferentialist program abdicates the responsibility of explaining the productive nature of *some* languages. Quite the contrary; Brandom takes his account of subsentential meanings to constitute one of the core achievements of the inferentialist program.

The primary notion of an inferentialist semantics for subsentential expressions is that of *substitution*, which Brandom inherits from (a reconstructed time-slice of) Frege. Starting with a finite stock of sentence types, Σ, each of whose free-standing (i.e., unembedded) uses have the default pragmatic significance of performing an assertion, we can ask whether any members of Σ can be treated as *substitutional variants* of any others. Keeping to the level of naïve intuition, the sentence 'David admires Herbie' is a substitutional variant of 'Jessica admires Herbie'. We'll see more about how this works in a moment, but the key takeaway point is this: if a sentence has a set of substitutional variants, then we can, *to that extent*, discern its subsentential structure. That is, by relating one sentence to another inferentially via substitution, we can notice and distinguish re-combinable subsentential expressions *within* the sentences of the language. Let's work through an example.

Take the sentence 'David admires Herbie' and chop it up any way you like, in respect of phonology, orthography, or whatever surface-level features happen to be relevant to the language at hand.[11] One way of doing so will yield 'Herbie' as a proper part; another yields '…

[10] For a dissenting view, see Stainton (2006).

[11] We'll do things in terms of orthography here (given the medium), but phonology is plainly the more primitive of the two in the human case, both phylo- and ontogenetically, as textbooks in empirical linguistics have long emphasized. For future robots, the medium will likely be something else—perhaps some descendent of TCP/IP. This would require adapting the substitutional techniques to that particular case.

…'erbi'...'. Now do the same with every other member of Σ, where the latter is assumed to be finite.[12] This yields a set of subsentential bits, Γ, consisting largely of nonsense like …'dmire'… and …'vid admi'…. With this in hand, go back to 'David admires Herbie' and substitute any other member of Γ (or, for that matter, Σ) in place of 'Herbie'. You'll find that most such substitutions yield uninterpretable gibberish—i.e., expressions that can enter into no inferential relations with the antecedently interpreted members of Σ. For instance, substituting 'jump' for 'Herbie' yields 'David admires jump', which has no inferential consequences. Same for 'jumps rapidly', 'red', 'we', and …'rential cons'…. By contrast, a commitment to 'Colorless green ideas sleep furiously' would presumably preclude entitlement to 'Nothing ever sleeps furiously', 'There are no colorless green things', 'Ideas can only be red', and many other propositions. There is a clear sense, then, in which this famous sentence is perfectly well interpretable. (It's even false!)

Setting aside gibberish, there will be a subclass of expressions that, when substituted for 'Herbie' in 'David admires Herbie', yield *interpretable* sentences, such as 'David admires Jessica', 'Jessica admires Herbie', 'David feeds Herbie', and 'David feeds Jessica'. (Again, a sentence is interpretable just in case it can play the role of premise or conclusion in an inference.) This subclass of Γ, call it Π, contains all and only the recombinable elements—i.e., the subsentential units of the language—including words, phrases, clauses, morphemes, subjects, predicates, or whatever other syntactic categories the language in question contains. We can now call one sentence, S, a *substitutional variant* of another, S*, just in case S is the result of substituting one element of Π with another in any member of Σ. Thus, 'David admires Herbie' is a substitutional variant of 'Jessica admires Herbie', on account of its being the result of the substitution of 'David' for 'Jessica'.

The foregoing puts us in a position to entertain a new inferential relation between sentences. Let's call an inference *substitutional* just in case the conclusion is a substitutional variant of one of the premises. The two inferences, from 'David admires Herbie' to either 'David feeds Herbie' or to 'David admires Jessica', are both fine examples. This no-

[12] Any *actual* creature's primary linguistic data (PLD) will, of necessity, be finite for in the course of language acquisition. The obvious analogy to the case of language acquisition in *human* children should not tempt us into assuming that Brandom is pitching an empirical account of the stages of acquisition. Still, the analogy is worth noting, even if we strongly suspect—as generativists do, *pace* Tomasello (2005)—that children's linguistic capacities are productive/generative *right from the get-go*. From the latter hypothesis, it follows that there is no such thing, really, as a finite set of PLD for the child, the child's acquisition device is *always* doing something analogous to hypothesis testing, *even in the absence of input data*. On this picture, the set of PLD is a constantly-moving target—in effect, a massively complex mental representation, or representational structure/system, within the child. The latter is plainly *not* identical with the set of utterances that happened to be produced in a child's presence.

tion of a substitutional inference is what allows for an application of the inferentialist strategy to subsentential expressions.

Subsentential Meaning: The meaning of a subsentential expression, α, is the set of *materially good substitution inferences* involving α.

Thus, the meaning of 'Herbie' is the set of inferential proprieties that includes {'David feeds Herbie' ↔ 'David feeds Herb'}, {'David feeds Herbie' ↔ 'David feeds a dog}, and {'David feeds Herbie' ↔ 'David feeds *his* dog'}, and many others.[13] In all such cases, the substitutional inferences are *materially good* in virtue of the fact that 'Herbie' is substituted by any of his other actual names, or by other ways of correctly describing him, uniquely or otherwise.

Needless to say, no one—not I, and certainly not Herbie—will ever have a *full* grasp of the set of inferential proprieties that governs the use of the expression 'Herbie', as this would involve knowing everything there is to know about him. Nor is there any guarantee that any two speakers will converge from the outset on what is correct to infer from "David feeds Herbie"—e.g., whether inferring "David feeds a dog" is (materially) good. Rather, the point is this: given that there are, in point of fact, *plenty* of ways for me to entitle myself to "Herbie is a dog", and no plausible ways (please grant) to undercut that entitlement, it would be *incorrect*, pragmatically *improper*, and epistemically *unwarranted* for someone to assert the opposite. This holds even if my interlocutor is *strongly disposed* to maintain a contrary position on the matter (foolishly, no doubt). It's important to always keep in mind that *normative* inferentialism is not about inferential *propensities*; it's about inferential *proprieties*.

1.6. *Predicates and singular terms*

One consequence of the view presented thus far is that some linguistic expressions can be inferentially stronger or weaker than others. Consider the verbs 'runs' and 'moves'. The latter is logically stronger than the former because all substitution inferences from '*x* runs' to '*x* moves' are good, but the reverse inferences generally aren't. In such cases, the substitution inferences are said to be *asymmetric*. We also find terms that *invariably* enter into *symmetric* substitution inferences—e.g., from 'Mark Twain was an American' to 'Samuel Clemens was an American' and back again. To make the latter type of inference explicit, subsentential expressions of identity and nonidentity can be introduced, yielding propositions of the form $\alpha=\beta$ and $\alpha\neq\beta$ (e.g., 'Sam Clemens is identical with Mark Twain' and 'David is not Herbie').[14]

[13] For simplicity of presentation, I suppress issues to do with possessives like 'my' and 'his', and indexical expressions more generally. Brandom (1994, 2008) supplies an account of these, but the details are irrelevant here.

[14] The notion of "introduction" that I intend here is the one developed in Brandom (2008). Roughly, a community is capable of introducing a novel expression,

As we will see in §4, Brandom holds that the distinction between predicates and singular terms comes down to the distinction between those expressions that *must* license *only* symmetric inferences (e.g., 'Herbie' and 'the dog'), and those that merely *can* license symmetric inferences, but *need not* do so (e.g., 'deer-like', 'jumps', and 'rapidly). On the basis of this claim, Brandom goes on to develop a complex line of reasoning whose ultimate conclusion I'll call the "asymmetry constraint".

Asymmetry Constraint: Any language that draws no distinction between predicates and singular terms (conceived in the above manner) *is in principle precluded from introducing conditionals*—i.e., expressions that make explicit one's commitment to the goodness of an inference—and other basic operators of propositional logic.

This claim will come to the foreground when we contrast it with Pietroski's predicativism, according to which there are in fact no singular terms at all in natural languages. If Brandom's argument succeeds, then Pietroski's predicativist semantic theory faces a serious challenge. Contrapositively, if Pietroski's predicativism is correct, then there must be a flaw in Brandom's reasoning. This is, in fact, the final puzzle for the overall reconciliation project that I'll be urging here.

1.7. *Types, tokens, and anaphoric chains*

The expressions discussed thus far have all been linguistic *types*, tokens of which may well diverge in meaning from their primary significance in the language. Indeed, terms like 'Herbie' have so many different uses—one for my dog, another for the pianist, Herbie Hancock, and countless others—that Brandom needs an account of what makes any use of 'Herbie' *semantically co-typical* with any other. The question applies even to *intra*-sentential occurrences: What makes it the case, for instance, that both tokens of 'Herbie' in 'Herbie admires Herbie' of the same type in a given communicative context?

In providing his answer, Brandom introduces the last of the major technical notions that he needs in order to carry off his overall project—viz., the notion of *anaphora*. Linguists and philosophers have paid a great deal of attention to *intra*-sentential anaphora, as in 'If a man is a

in this sense, just in case its members already have the *practical* abilities that are necessary and sufficient for being able to express—i.e., to make explicit—normative attitudes that were previously only implicit in their practice. Thus, the practical ability to implicitly treat someone as having entitled themselves to q by committing themselves to p is both necessary and sufficient to introduce conditional expressions that make explicit the material goodness of that inference—e.g., '$p \supset q$' and 'If p then q'. We will see in §2.5 that Pietroski's notion of concept introduction is different from Brandom's, and arguably orthogonal.

police officer, then he was born out of wedlock', where the pronoun 'he' is anaphoric on 'a man'. Syntacticians, in particular, have devised principles of generative grammar that aim to explain the natural distribution of anaphoric expressions within sentences of natural language. Somewhat less effort has thus far been expended on analyzing *inter*-sentential anaphora, as in the following exchange between speakers Mihir and Rushal.

> Mihir: That man seems to have fallen ill right after he approached the police line.
> Rushal: He must have gotten hit by their fancy new sonic weapon.
> Mihir: Oh, hey, I didn't see you there! Do you happen to know the guy?
> Rushal: No, I just heard you talking about him and I figured I'd chime in.

Here, an anaphoric chain is initiated by Mihir's use of 'That man', which is then picked up by 'he' later in the same sentence. But the chain doesn't end there. Rushal's use of 'He' is anaphoric on Mihir's use of 'That man' and 'he'. Mihir's response picks up the anaphoric chain with an occurrence of 'the guy', which then continues onward to Rushal's use of 'him', and to occurrences of other expressions in subsequent discourse. Setting aside syntactic issues, what can we say about this phenomenon at the level of *meaning*?

In keeping with his inferentialist semantics, Brandom argues that an anaphoric chain is one in which the inferential proprieties governing the anaphoric initiator (e.g., Mihir's use of 'That man') are *inherited* by subsequent expressions in the chain. Thus, if Mihir's use of 'That man' is partly governed by his commitment to 'That man is falling ill on live television', then Rushal inherits this commitment (among others) in picking up the anaphoric chain with the use of 'He', along with whatever entitlements for this claim Mihir had already secured prior to Rushal's appearance on the scene.

With this account in hand, Brandom treats as a special case occurrences that are treated as semantically co-typical *because* they are phonologically or orthographically co-typical—e.g., the two occurrences of 'Herbie' in 'Herbie admires Herbie'. From this perspective, *all* expression types consist of long-stretching anaphoric chains of individual use—an idea familiar from causal theories of reference-borrowing, though shorn of various optional commitments. This account also makes it clear what's happening at the level of *pragmatics*. In picking up anaphoric chains, speakers are able to take on normative statuses—paradigmatically, commitments and entitlements—without themselves having *explicitly avowed* those statuses, and often without having much (if any) idea what exactly it is that they've inherited. To illustrate, we can extend the above example.

Suppose Rushal had no prior commitments regarding the victim's appearance on television, or indeed anything at all about the victim, but was strongly committed to the claim that police don't use sonic weapons on camera. In that case, upon being subsequently apprised of Mihir's entitlement to 'That man is falling ill on live television', Rushal will be under normative pressure to either revise his prior commitments about on-camera police violence, or to withdraw the claim that the victim *must* have been affected by a specifically *sonic* weapon. In this second case, the revision can target either the predicate 'sonic'—perhaps the police used an invisible gas—or the alethic modal expression ,'must'. The latter, on Brandom's view, functions to make explicit the *modal robustness* of an inference—i.e., its insensitivity to substitutions of background auxiliary commitments, up to some boundary conditions (e.g., physical law). In the present case, the boundary conditions are set by Rushal's commitments regarding the general institutional practices of local police. In order to regain epistemic equilibrium, Rushal can revise various commitments concerning these practices; for instance, he might conclude that the local Sheriff has deemed this to be a special occasion, on which on-camera use of sonic weapons is warranted.

1.8. *Summary*

We've now surveyed the main contours of Brandom's overall philosophical project. The explanatory strategy he pursues can be characterized as "top-down", in the sense that he begins by offering an account of communal normative practices, in the broadest sense, and identifies within these an important subclass—namely, practices that serve to institute distinctively *linguistic* norms governing assertion and other communicative acts. (One last plea for demands!) Such norms pertain to the inferential proprieties that expression types have in their semantically primary occurrences. Thus, the account moves "down" a step—from a normative pragmatics that posits statuses of commitment and entitlement, to an inferentialist semantics that aims to analyze *meaning* in *terms of* these statuses. The meaning of a propositional expression type is, on this picture, identified with its normative inferential role—i.e., what *other* claims it commits or entitles one to, and what commitments one must undertake in order secure an entitlement to it.

Drilling down still further, Brandom develops the substitutional approach, which allows one to "dissect" proposition-sized expression types, revealing subsentential bits of vocabulary. These carry their own "ingredient content", despite lacking the *free-standing* significance of propositional expressions that enter directly into inferences as premises or conclusions. The details of this proposal put in place the theoretical commitments that Brandom needs in order to distinguish predicates from singular terms—a distinction that he goes on to argue will be discernable in *any* linguistic practice that allows for the introduction of conditionals and other logical operators (§4.3).

Having offered a treatment of propositional and subsentential expression types, Brandom steps down another rung on the explanatory ladder, developing a conception of *anaphora* that applies far more broadly than standard discussions in the literature might lead one to suspect. The anaphoric relationship is, on this view, one of *inferential inheritance*, wherein the proprieties governing the use of one expression—the initiator of an anaphoric chain—are taken to then *also* govern the expressions occurring later in the chain, irrespective of the speaker's acknowledgement (or even awareness) of the statuses they've thereby undertaken. The latter condition serves to explain how speakers can felicitously use expressions whose total set of inferential proprieties is *unknown* to them, and perhaps even to *anyone* in the community.

One might think that all of this is utterly wrongheaded right from the get-go—the normativity, the substitutions, and even the top-level goal of delineating language-use *as such*. Indeed, from the perspective of a mainstream contemporary linguist or philosopher of language, Brandom's whole "top-down" explanatory strategy will seem downright perverse. The more common *bottom-up* alternative goes as follows.

Taking for granted the notions of denotation/reference and satisfaction, as applied to subsentential expressions, the bottom-up theorist seeks to formalize a compositional apparatus for building propositions out of them. Free-standing propositional complexes are thereby recursively assigned their own special kind of semantic value: e.g., possible-worlds truth conditions (Heim and Kratzer 1998) or sets of possible worlds (Stalnaker 1984). This, in turn, opens the door to a theory of linguistic *communication*, according to which speakers append illocutionary forces to the range of recursively-specified meanings, yielding a variety of speech-act types (questions, commands, etc.). The inferences in which a (now-interpreted) speech act type figures can then be classified as good or bad in virtue of the semantic structures that the combinatorial apparatus assigns to their premises and conclusions, as well as the illocutionary forces that (somehow) "attach" to those structures.

Having thus analyzed the semantic properties of speech acts and inferences, one might note that some—perhaps, in the end, *all*—of these have features that reliably trigger unencapsulated pragmatic reasoning. This motivates the familiar project of supplementing a pragmatic theory with "maxims" of rational cooperative communication/action (Grice 1989; Sperber and Wilson, 1986). Theorists who have carried out this latter project have developed impressive accounts of implicature, metaphor, and other complex communicative phenomena (Levinson 1983; Harris 2020).

Proponents of the bottom-up strategy have pressed a catalogue of objections to Brandom's project. These include, but are not limited to, the following: (i) insistence on a *compositionality* constraint that the inferentialist allegedly can't accommodate; (ii) rejection of the idea that language is fundamentally a *communicative* system; (iii) requirement

that any *legitimate* inquiry foreswear trafficking in normative assessments; and (iv) an allegation to the effect that normative inferentialism is incompatible with what is known empirically about the human mind/brain, particularly in respect of its language-processing abilities.

Before any of these challenges can be met, each stands in need of careful articulation. As previously noted, I believe that such a task is best undertaken by pitting Brandom's project against what appears, at first blush, to be a rival alternative. (As advertised, I'll argue afterwards that the appearances are often deceiving in this regard.) With that in mind, I now turn to the work of Paul Pietroski, whose semantic theory is a recent and powerful contribution to the larger enterprise of generative linguistics.

2. *Pietroski: Meanings as pronounceable instructions for concept assembly*

The theoretical commitments that comprise Paul Pietroski's approach to natural-language semantics are advanced and defended in his recent book, *Conjoining Meanings* (henceforth *CM*).[15] In this section, I summarize several of Pietroski's main contributions, highlighting aspects of his view that bear on my ecumenical strategy in §§3-4. To be clear from the outset, the ideas laid out in *CM* strike me as constituting genuine progress in our understanding of the psychological mechanisms of human language use. Moreover, I find wholly compelling his arguments against the central pillars of received views in semantics—particularly, the commitment to an extensional/truth-conditional approach. The book, overall, is replete with rich and instructive discussions of topics that go well beyond the scope of the present discussion. But while we won't be able to look at the details of some of Pietroski's original proposals here, it's worth noting that they are all, to my mind, persuasively motivated by historical, formal, and empirical considerations. That having been said, let's dive in.

2.1. *A different methodology and new explanatory aims*

While Brandom's inferentialist approach is virtually unknown in cognitive science, the methodology of generative linguistics will be familiar to many in the field, at least in broad outline. Rooted in a foundational commitment to *naturalistic* inquiry, the idea is to treat language as a biological phenomenon—not necessarily in the sense that it has an adaptive function (Chomsky [2016] disputes this), but in the sense that a neurophysiologically realized cognitive structure is the explicit target of inquiry. The linguist thus works on the assumption that human minds contain a language-specific device—a "faculty", "module", or "mental organ"—with a distinctive computational architecture, a

[15] This section elaborates the material in Pereplyotchik (2019). The operative notion of a *subpersonal* level of description is spelled out in Pereplyotchik (2017: ch. 7).

proprietary representational format, and dedicated/domain-specific information-processing routines. The goal is to provide a detailed specification of each of these, yielding a neurocognitive account of the acquisition and use of language.

On analogy with bodily organs, the faculty of language (henceforth FL) is assumed to "grow" within the child during the early years of development. This happens in accordance with a genetic program, phenotypically realized in the child's innate ability to acquire linguistic competence under a diverse range of social and environmental circumstances. Thus, a central aim of generative linguistics is to specify not only the grammar of an adult language, but also the principles that underlie language *acquisition*—particularly those that allow the child to home in on a *specific* grammar in a relatively short time, with little or no (overt) negative evidence (Chomsky 1986; Yang 2006). This problem is made exceedingly challenging by the fact that natural languages are invariably productive/generative, meaning that they allow for boundless applications of combinatorial recursive operations, yielding a discrete infinity of nonredundant[16] interpretable structures.

The generativist's strategy for dealing with this central feature of natural language is to posit grammatical principles that are inherently *compositional* at all levels of analysis—phonology, morphology, syntax, and semantics. The syntactic module of FL is taken to merge the elements of the lexicon—atomic units of a language that contribute their distinctive meanings to more complex structures. On the basis of these, the semantic module recursively generates complex *meanings*, which can enter into downstream personal-level cognition—judgment, reasoning, planning, and the like.[17]

Pietroski's main goal in CM is to characterize the semantic module by offering a detailed proposal about its proprietary representational format—specifically, the nature of the lexical items—and the computational operations that assemble larger interpretable structures. At the level of format, the hypothesis he develops is that virtually all lexical items are predicates, the latter being restricted to only two types—monadic and (semi)dyadic. Regarding computational operations, Pietroski aims to make do with a bare minimum of compositional semantic principles, with the lion's share of work being done by nothing more than two flavors of predicate conjunction (one for each type of predicate).

[16] Pietroski points out that this goes well beyond mere recursion, which is trivially satisfied by any languages with a rule for applying sentential operators. The infinitude of English thus differs *qualitatively* from the infinitude of a language that permits the formation only of P, P&P, P&(P&P),… or P, ~P, ~~P, ~~~P,….

[17] It's important to note that what has been said thus far is not (yet) intended as a theory of real-time/on-line language processing. Rather, it is to be seen as an abstract characterization of the architecture and internal operations of a specific cognitive structure, acquired at birth and persisting in a stable state thereafter (Chomsky 1995).

We'll look at some of the details shortly, maintaining our present focus on matters of methodology.

Following Chomsky (1986, 1995, 2000), Pietroski adopts an *individualist* position, taking the object of study to be an "I-language"—an *i*ntensionally-specified procedure *i*nternal to an *i*ndividual language user. He supports this with forceful arguments against the alternative conception of language(s) that we find in the work of David Lewis (1969, 1970, 1973). A language, on this rival picture, is a kind of abstract object—namely, an extensionally-specified a set of well-formed sentences—which is "selected" by a population of creatures, via the adoption of social/communicative "conventions". The latter Lewis sees as jointly constituting a *public* language, such as English or Norwegian—what generativists refer to as "E-languages". Pietroski rejects virtually every aspect of this picture. We'll look at his reasons for doing so in §4. For now, it's sufficient to distinguish three key points of contention.

First, there's the metaphysics. Lewis (1973) says languages are abstracta, whereas Pietroski sees them as biologically-instantiated computational procedures. Then there's the issue of extensionality. Pietroski rejects Lewis's theoretical goals, which consist merely of extensionally specifying meaning-pronunciation pairs, and adopts instead a more weighty explanatory aim—namely, that of specifying human linguistic competence as a function-in-*intension*. Only in this way, he argues, can the resulting theory capture the *psychologically real* operations that yield interpretable structures. Finally, there's the issue of publicity, and related troubles with Lewis's notion of "selection". Pietroski's individualist stance leads him to eschew the folk-ontological commitment to public languages, at least for the purposes of mature empirical inquiry. This manifests in his methodological practice of focusing on matters of individual psychology—e.g., internal mechanisms of semantic composition—rather than the social practices of linguistic communication. Accordingly, Pietroski sees Lewis's appeal to public conventions as generally unhelpful for—indeed, an outright *distraction from*—the empirical study of linguistic meaning.

Pietroski's disagreements with Lewis go well beyond such methodological issues, extending to matters of technical detail. For, in addition to the large-scale commitments mentioned thus far, Lewis (1970) also developed a powerful formal apparatus for conducting semantic theorizing. Expressions, in this scheme, are assigned "semantic types", which are either basic or recursively derived. The interpretation of complex structures is then accomplished by functions that map one semantic type onto another. In its most familiar version, such a semantic theory will assign sentences the basic type <t> and singular terms the basic type <e>. Thereafter, monadic predicates can be treated as having the derived type <<e>, <t>>, which is a function from things of type <e> to things of type <t>.

Although this formal typology presupposes no particular metaphysics or metasemantics, it's common in practice to think of singular terms

as denoting entities (e.g., Jessica), and sentences as denoting truth-values (T and F). With this in place, monadic predicates like 'swims' can be assigned the semantic function of mapping the entities in its domain to the truth-values in its range. For instance, 'Jessica swims' is mapped to T just in case Jessica (the actual person) satisfies the predicate 'swims'; otherwise, F. Likewise, adverbs such as 'often' and 'expertly' have the derived type <<<e>, <t>>, <t>>, which is a function that maps the semantic value of predicates (i.e., functions from <e> to <t>) to the semantic values of sentences (i.e., T or F). Put somewhat imprecisely, the intuitive idea is that 'Jessica swims expertly' is mapped to T just in case 'expertly' is satisfied by the predicate 'swims' when applied to 'Jessica'.

It's no exaggeration to say that this general framework is seen as a foundational contribution to formal semantics, even by generative linguists who have no truck with—or, indeed, no awareness of—Lewis's broader projects. Part of what makes Pietroski's negative contentions so radical, then, is that he rejects wholesale this now-mainstream approach to semantic theorizing. In particular, he argues that taking an infinite hierarchy of types as explanatorily primitive is not only unparsimonious, but leaves wholly unexplained crucial aspects of the natural languages that children invariably acquire. As a matter of empirical fact, humans language permits the construction of only a *limited* class of semantic types, not the infinite range of logically possible ones. This empirical generalization plainly stands in need of explanation, which a semantic theory can't provide if it takes all possible types as available to a speaker right from the start.

> One can say that thinkers must have the requisite abstractive powers, given the capacities required to form thoughts like ABOVE(FIDO, VENUS) & BETWEEN (SADIE, BESSIE, VENUS). But one needs an account of these alleged powers—which permit abstraction of a tetradic concept from ABOVE(_, _) and BETWEEN(_, _, _)—to explain how thinkers can form the concepts that *Begriffsschift* expressions reflect. This is not to doubt the utility of Frege's logical syntax. On the contrary, his proposals about the architecture of thoughts were major contributions. But Frege insightfully invented a logical syntax whose intended interpretation raised important questions that he did not answer.
>
> One can insist that given any polyadic concept with n unsaturated "slots," a human thinker can use n-1 saturaters to create a monadic concept, leaving any one of the slots unfilled. But that leaves the question of how we came to have this impressive capacity. And in chapter six, I offer evidence that a simple form of conjunction lies at the core of unbounded cognitive productivity. Our natural capacities to combine concepts are impressive, but constrained in ways that suggest less than an ideal Fregean mind.

Pietroski recommends a more parsimonious alternative—one that eschews the infinite hierarchy of semantic types and posits only a very small handful, including, most importantly, monadic and quasi-dyadic predicates. "The idea [is] that with help from Slang syntax, we can generate an analog of GIVE(VENUS, _, BESSIE) without saturating GIVE(_, _, _)

—much less saturating it twice, or thrice, and then desaturating once" (103).

Nor does his iconoclasm end there. As noted earlier, Lewis's general framework for semantic theorizing leaves open a variety of issues in metaphysics and metasemantics. An equally mainstream approach to natural-language semantics is decidedly more committal on these points. Donald Davidson's truth-theoretic semantics (Davidson, 1983), as well as the many variants of it that have now been developed, identifies the meanings of linguistic expressions with their *extensions*. Thus, truth conditions (perhaps relativized to possible worlds) are seen as the semantic values of sentences; entities are the values of singular terms; sets are the values of predicates; events in the case of verbs, and so on. Pietroski marshals a battery of arguments against this familiar approach. We'll examine these shortly. For now, we note only that this anti-extensionalism is a core commitment that he shares with Brandom. It is, therefore, a major plank in the bridge that I aim to build between the two in §§3-4.

2.2. *Meanings are definitely not extensions*

Pietroski sees semantics as a naturalistic inquiry into "how Slang expressions are related to human concepts" (115). Some theorists wish to simply *identify* meanings with concepts, but Pietroski points out that this leaves wholly unexplained the psychological processes that *constitute* our semantic competence. I'll argue in §4 that this point applies to Brandom, who sometimes speaks indiscriminately of meanings, concepts, conceptual contents, intentional contents, discursive contents, propositional contents, and so on. However, as I'll emphasize there, the difference can only be viewed as a substantive theoretical *dispute* if we let their use of the folk term 'meaning' bewitch us into assuming that they have a common explanatory target, contrary to fact.

Better, I think, to appreciate the highly theoretical nature of this piece of jargon and the different—*but not thereby incompatible*—explanatory goals of the two frameworks in which it shows up. Thus, we can distinguish meaning_B from meaning_P and proceed to contemplate how the two are related, this now being a jointly philosophical and empirical question, not a boring verbal one. Indeed, this point is made explicitly by both Pietroski and Brandom, in connection with both 'meaning' and another vexed notion—that of 'concepts'—which notoriously plays a wide variety of roles in diverse research contexts. Here again, we can speak of concepts_B and concepts_P, aiming to articulate the relations between them. Likewise for 'thought', 'judgment', and other terms, when explicit disambiguation is required. (See also the discussion of the notorious '-ing/-ed' ambiguity in §4.2.)

As noted above, another popular idea is to identify meanings with extensions (Davidson 1983). The central negative contention of *CM* is that the notions of extension, truth, and denotation should play *no ex-*

planatory role in a psychologically-oriented semantics for natural languages ("Slangs"). Pietroski argues persuasively that the best empirical theory of the relation between Slang expressions and concepts will *not* identify meanings with extensions. Indeed, he rejects even the weaker claim that meanings *determine* extensions. He proposes, instead, to identify meanings with something entirely different—in particular, something that can play the psychological role of relating language to cognition. The candidate he recommends is this: *pronounceable instructions for accessing and assembling concepts*. We'll look at this in some detail, but let's first get clear on *why* Pietroski rejects the truth-conditional orthodoxy that dominates formal semantics. As we'll see, there are a great many reasons. To my mind, no one of these is necessarily decisive, but, taken together, they strongly suggest turning away from the extensionalist project and starting anew, *however* much revision this might require. As we go along, I'll land a few jabs of my own.

2.3. *Objections to truth-conditional semantics*

Pietroski views truth-conditional semantics (henceforth 'TCS') as an empirical hypothesis about Slang expressions, according to which there is a relation—call it "true of", "refers to", "denotes", or whatever you like—that holds between words and items in the world. TCS views this relation as being of central importance to our theoretical characterization of natural-language meanings. In rejecting this hypothesis, one need not deny, of course, that there are words or that there objects (e.g., babies and 'bathwater'). One can, instead, deny that there is a unique relation between them, let alone one that's suited to playing the theoretical role of *linguistic meaning*. Here is how Pietroski puts the point:

> I don't think 'sky' is true of skies (or of sky), much less blue skies or night skies. I don't deny that there are chases, and that in this sense, chases exist even if skies don't. But the existence of chases doesn't show that 'chase' is true of them… [Likewise], there is no entity that 'Venice' denotes. In this respect, 'Venice' is like 'Vulcan', even though one can visit Venice but not Vulcan… I also agree that there is a sense in which there are blue skies, but no blue unicorns. But it doesn't follow that 'sky' is true of some things, at least not in the sense of 'true' that matters for a theory of truth… [T]here is no call to quantify over skies, in physics or linguistics. (68)

As the example of 'Vulcan' illustrates, words can perfectly well be meaningful without having extensions. Pietroski's view is that this holds of *all* Slang expressions. What's interesting about words like 'Vulcan' is they "illustrate the general point that words don't *have* extensions". The idea isn't merely such terms have *empty* extensions; it's that they have *none at all*.

Even if words *did* have extensions, the latter couldn't be identified with meanings, if only because "expressions with different meanings can have the same 'extension'" (15). Fans of TCS will typically appeal to "non-actual possibilities" in dealing with this issue. For instance, 'unicorn' and 'ghost' are said to have the same extension in the *actual*

world, but they differ in meaning—the reply goes—because they have *different* extensions in *other* possible worlds. Pietroski correctly points out that this "is an odd way to maintain that meanings are extensions."

> If the meaning of a word is not whatever set of things that the word happens to be true of, why think the meaning is a mapping from each possible world w to whatever set of things that the word happens to be true of at w? [If] Slang expressions need not connect pronunciations to actual things, it seems contrived to insist that these expressions connect pronunciations to possible things… [I]nvoking possible unicorns is contrivance on stilts. (12)

Doubtless, fans of TCS will see this as little more than an *ad hominem*. We'll look at stronger arguments shortly. For now, I want to emphasize that this point—or, in any case, a version of it—carries more weight than is commonly appreciated. Let me take a brief aside to develop it in my own terms.

The intuitive considerations that motivate TCS (e.g., for introductory semantics students) almost always have to do with objects that are available for perceptual inspection. ('David' refers to *this guy*, 'my desk' refers to *that thing*, and so on.) This serves to illustrate, at the level of pre-theoretical intuition, how linguistic expressions "hook onto the world"—namely by way of perceptual contact (indeed, literal contact, in the case of haptic perception). Shortly thereafter, the details of one or another formal theory are introduced, giving the student little time to reflect on how far the initial illustration can plausibly generalize. (Spoiler alert: not very far!) If philosophical questions happen to arise about the status of these "reference" and "correspondence" "relations"—e.g., with regard to empty names and predicates ('Vulcan', 'unicorn', etc.)—the instructor can use the opportunity to explore various technical proposals for dealing with such "special cases"—e.g., Russell's theory of names as disguised descriptions, or the formal apparatus of possible-worlds semantics. Attention is thus deflected away from how massive the intuitive problem really is. Here's a much-needed corrective.

Consider for a moment the vast range of expressions that we can readily produce and comprehend, and reflect on how vanishingly few of these have anything much to do with what's going on in physical reality, let alone with things that we can perceptually inspect in any intuitive sense. We speak of Santa and his elves, gods and demons, goals and fears, opportunities and temptations, aliens and chem-trails, reptiles and unicorns, futures and fictions, numbers and functions, nouns and verbs, fonts and meanings, haircuts and field-goals, stocks and derivatives, mergers and monopolies, economies and governments, boson fields and spin-foams, black holes and electrons, Blacks and whites, Jews and Frenchmen, London and Moscow, classes and genders, protests and stereotypes, jocks and nerds, bits and bytes, poems and operas, humor and beauty, and even the possibility (albeit dim) of true liberatory justice.

Appreciating the sheer scope of the phenomenon to be explained renders, to my mind, *utterly implausible* the strategy of taking direct perceptual contact with the world as our model of how language relates to reality *in general*. Moreover, the total lack of convergence that we find amongst metasemanticists when we go looking for a metaphysical account of truth and reference—conceived of, again, as a Very Special sort of natural relation—strikes me as further grounds for abandoning the project of extensional semantics immediately and forthwith. It helps, of course, that Pietroski supplies a powerful *alternative* framework for doing semantics. And it certainly doesn't hurt, that Brandom complements this with an independently attractive ("deflationist") account of truth and reference.

All that aside, Pietroski has a further, more powerful argument against invoking non-actual possibilities for the purpose of individuating meanings. He makes use of Kripke's contention that the non-existence of unicorns in the actual world implies their non-existence in *all other* possible worlds (Kripke, 1980). Of course, there may well be creatures in other possible worlds that *look* a lot like what we imagine unicorns would look like. But they would not thereby *be* unicorns, and our word 'unicorn' would not thereby be true of them. If that's correct, then 'ghost' and 'unicorn' aren't just co-extensive in *our* world; they're co-extensive in *every* possible world. Thus, no identification of meanings with extensions, actual or possible, will distinguish the meanings of those two words. Likewise for all of the related cases—empty names, defective predicates, necessary falsehoods, and so on.

One might reply by rejecting Kripke's semantic and metaphysical assumptions, and adopting instead a Lewisian counterpart theory, but Pietroski points out several problems for this strategy as well. Adopting the terms 'LUNICORN' for Lewisian unicorn-lookalikes and 'KUNICORN' for the whatever it is that Kripke has in mind, he makes the following powerful retort.

> We can grant that some theorists sometimes use 'unicorn' to express the technical concept LUNICORN. But if 'unicorn' can also be used to express the concept KUNICORN, then it seems like contrivance to insist that the Slang expression has a meaning that maps some contexts onto the extension of LUNICORN and other contexts onto the extension of KUNICORN. If we assume that words like 'possibly' have extensions, then perhaps we should specify the meanings of such words in terms of a suitably generic notion of world that allows for special cases corresponding to metaphysical and epistemic modalities; cp. Kratzer. But in my view, theorists should not posit (things that include) unicorns in order to accommodate correct uses of 'Possibly/Perhaps/Maybe unicorns exist' or 'There may be unicorns'; and likewise for squarable circles.

Thereafter, the dialectic turns to matters that we need not enter into here. Suffice it to say that, even if this worry about fine-grained meanings can ultimately be defused, TCS would still face Pietroski's more technical (and potentially more damaging) objections. These include

matters pertaining to liar-sentences, as well as the more widespread and natural phenomenon of event descriptions. These too go beyond the scope of our discussion. One argument that I do want to say a bit more about, though, is on the topic of polysemy, where Pietroski's view of the matter finds wide acceptance among generative linguists—though, notably, not philosophers of language (see, e.g., Michael Devitt's paper in this issue.)

Following Chomsky (2000), Pietroski points out that 'water' is polysemously used to talk about many substances—those found in wells, rivers, taps, etc.—nearly all of which have *lower* H_2O contents than substances that, at least *prima facie*, are *not* water, including coffee, tea, and cola (*CM*, 21). This presents a challenge to theories that view 'water' as bearing a reference relation to (all instances of?) the natural kind *water*, whose metaphysically essential property is *being composed of H_2O molecules* (Kripke, 1980). If coffee, tea, and cola all have more H_2O in them than most ordinary instances of water, then it's not clear why 'water' doesn't bear the reference relation to *them*, rather to the stuff in the local rivers and wells.

A related consideration has to do with predicate conjunction. The word 'France' can be used in expressing either of two concepts: FRANCE:BORDER and FRANCE:POLIS. The border is hexagonal and the polis is a republic. But, Pietroski points out, the polysemy of 'France' "does not imply that something is both hexagonal *and* a republic, much less that 'France' denotes such a thing" (74). Similarly, while 'London' can be used to talk about "a particular location or a polis that could be relocated elsewhere," it is plain that "no location can be moved, and no political institution is a location." Pietroski concludes that "no entity is the denotation (or 'semantic value') of 'London'; the ordinary word *has* no denotation" (73, emphasis mine).

2.4. *Meanings as pronounceable instructions*

Let's turn now to Pietroski's positive views. As noted earlier, the main goal of *CM* is to defend the hypothesis that linguistic meanings are "pronounceable instructions for how to access and assemble concepts" (1). More specifically,

> each lexical meaning is an instruction for how to access a concept from a certain address, which may be shared by a family of concepts. … A Slang expression Σ can be used to access/build/express a concept C that is less flexible than Σ—in terms of what Σ can be used to talk about, and how it can combine with other expressions, compared with what C can be used to think about and how it can combine with other concepts— since Σ might be used to access/build/express a related but distinct concept C* .

Unpacking Pietroski's hypothesis requires getting clear on the three key notions of *pronounceable instructions*, *compositional assembly*, and *conceptual types*. Each is more challenging than the last, so we'll start with instructions and work our way up.

2.4.1. *Pronounceable instructions*

An utterance of a sentence is a spatiotemporally located event, in which a speaker produces a physical signal. The latter serves, on Pietroski's view, as an instruction for the hearer's FL to perform a computational procedure.[18] The instruction can be carried out by any hearer whose I-language is sufficiently similar to the speaker's. The acoustic properties of an utterance, upon being transduced, trigger an early perceptual constancy effect, whereby a dedicated module imposes phonological categories on the neural encoding of the acoustic blast. These cognitive operations serve, in turn, as instructions for the further segmentation of the phonological units into syllables and eventually into morphemes and other lexical items. The latter, on Pietroski's view are best seen as instructions for accessing ("fetching") individual concepts, which he conceives of as atomic units of one or another language of thought. I say "one or another" because his view leaves open the possibility, which he goes on to explore and even endorse, that there are *many* languages in which the mind conducts its information-processing. We'll return to this point in connection with Pietroski's discussions of Frege (§2.5).

Importantly, Pietroski maintains that concepts reside in semantic "families", which have their own "addresses" in a broader cognitive architecture. This is a large part of his explanation of the aforementioned phenomenon of polysemy. The idea is that one and the same lexical item can be an instruction for fetching "a concept from a certain lexical address … shared by a *family* of concepts" (8). Because a lexical instruction points only at an address, rather than a specific concept, it's left open for *downstream* processing routines to determine which particular concept from the indicated address/family is "relevant" in the present context.

This, of course, raises deep and difficult questions about how hearers manage this latter step—i.e., reliably accessing the relevant concept(s) in a given context, rather than the irrelevant ones from the same conceptual family. What psychological mechanisms select just *one* of a family of concepts residing at a common lexical address? In large part, Pietroski leaves this issue open—justifiably so, given everything else he's juggling. But it's worth remarking in the present context that the mechanisms of this kind of selection are widely agreed to involve—indeed, to *require*—precisely the kind of nondemonstrative pragmatic reasoning that Brandom has argued to be constitutive of conceptual contents.

[18] "I hope the analogy to elementary computer programs, which can be compiled and implemented, makes the operative notion of instruction tolerably clear and unobjectionable in the present context. … Instructions can take many forms, including strings of '0's and '1's that get used—as in a von Neumann machine—to access other such strings and perform certain operations on them. … And instead of arithmetic operations that are performed on numbers, one can imagine combinatorial operations that are performed on concepts" (108).

2.4.2. *Assembling concepts*

Turn now to the second key notion in Pietroski's main hypothesis—viz., the compositional assembly of concepts. In general, instructions for assembling something can vary along any number of dimensions. Some are clear; some aren't. Some are detailed; others are vague. Some are simple; others are complex—i.e., composed of simpler instructions. Moreover, not everything to which an instruction is presented is capable of carrying it out. Some computers can't run the software that others can. Some chefs can't bake the cakes that others have no trouble baking. And some proteins (or cells) can follow genetic instructions that others simply can't. Lastly, the *products* of successfully carrying out instructions can vary widely. The same student, with the same instructions, can succeed or fail on an exam, depending on whether they've had sleep the night before. Likewise, a novice barista will generally make worse coffee with low-quality ingredients than with high-quality ones, successfully following the same instructions both times.

Given that the semantic module of FL is assumed to have a stable processing routines, carried out in a proprietary representational format, it follows that it won't be able to process just *any* old instruction, but only a restricted kind. Likewise, it will only be capable of assembling only a limited class of outputs. The question, then, is what kinds of instructions the semantic module is capable of implementing and what sort of structures it's capable of building.

Many theorists aim at capturing something called "compositionality"—a piece of theoretical jargon that, perhaps more than most, has been worn smooth by a thousand tongues (to use Wilfrid Sellars's clever phrase). Of the many ways of cashing it out, Pietroski maintains that what's required for an avowedly *cognitivist* project is that the meanings of lexical items compose in ways that suitably mirror the structure of complex concepts. Thus, having identified the meanings of lexical items with instructions to fetch individual concepts, he argues that these instructions compose, forming *complex* instructions, with some functioning as (detachable) components of others. These semantic instructions—what Pietroski calls *Begriffsplans*—are responsible for the assembly of concepts meet two constraints. First, they must be suited to *that specific type* of instruction. While other kinds of human concepts might be assembled by non-linguistic means, *Begriffsplans* can only assemble concepts of a very specific nature (to be spelled out shortly). Second, in keeping with the "mirroring" constraint (my word, not his), the complex concepts that *Begriffsplans* assemble must bear the same part-whole relationships to one another as do the *Begriffsplans* themselves.

Laying out some of the specifics of the *Begriffsplans* that Pietroski posits will put us in a position to better appreciate his views on concepts. The clearest case of this pertains to instructions for *predicate conjunction*. Pietroski takes this to be an absolutely central aspect of

linguistic concept assembly, in part because he holds that the kinds of concepts that the human FL is capable of assembling are uniformly *predicative*. In saying this, he means to deny outright that natural languages ("Slangs") allow us to access singular concepts. Such concepts do exist, he thinks, but they *can't* be fetched by *Begriffsplans*. Indeed, he holds that the *only* predicative concepts FL can fetch, and hence assemble, are limited to just the monadic and the quasi-dyadic, with higher adicities receiving a different analysis. These two types of concept correspond to two flavors of predicate conjunction: M-junction and D-junction. Here's how Pietroski characterizes the overall process.

> If biology somehow implements M-junction and D-junction, one can envision a mind with further capacities to (i) use lexical items as devices for accessing simple concepts that can be inputs to these operations, and (ii) combine lexical items in ways that invoke these operations. … Suppose that combining two Slang expressions, atomic or complex, is an instruction to send a pair of corresponding concepts to a "joiner" whose outputs can be inputs to further operations of joining. Imagine a mind—call it Joyce—that has some lexical items, each with a long-term address that may be shared by two or more polysemously related concepts. Joyce also has a workspace in which (copies of) two concepts can be either M-joined or D-joined to form a single monadic concept, thereby making room for another concept in the workspace, up to some limit. Joyce can produce and execute instructions like fetch@'cow'; where for each lexical item L, the instruction fetch@L is executed by copying a concept that resides at the long-term address of L into the workspace. Joyce can also produce and execute instructions of the forms M-join[I, I0] and D-join[I, I0]; where I and I0 are also generable instructions. An instance of M-join[I, I0] is executed by M-joining results of executing I and I0, leaving the result in the workspace, and likewise for an instance D-join[I, I0].

Having introduced two basic types of composable *Begriffsplans*—one for fetching concepts like DOG(_) and one for assembling these into complex structures—Pietroski adds four other types of basic semantic operation:

(i) a limited operation of *existential closure*
(ii) a mental analog of *relative clause formation* (weaker than λ-abstraction)
(iii) the *introduction* of concepts like GIVE(_) on the basis of GIVE(*x, y, z*)
(iv) of *thematic* concepts—e.g., AGENT(_), PATIENT(_), RECIPIENT(_)

> [G]iven two monadic concepts, the operation of M-junction yields a third such concept that applies to an entity e if and only if each of the two constituent concepts applies to e. (32) … In short, Slangs let us access and assemble monadic [and some limited dyadic] concepts that can be conjoined, indexed, polarized, and used as bases for a limited kind of abstraction.

We'll look at several of these operations in more detail below, but the following passage contains an initial illustration of the kinds of structures that this system can assemble.

> My claim is not that 'gave a dog a bone' is an instruction to build [just *any*] concept with which one can think about things that gave a dog a bone. That instruction might be executed by building the concept ∃y∃z[GAVE(x,

> y, z) & BONE(y) & DOG(z)], which has a triadic constituent. My claim is that 'gave a dog a bone' is an instruction for how to build an M-junction like [[GIVE(_)^PAST(_)]^∃[PATIENT(_, _), BONE(_)]]^∃[RECIPIENT (_, _) ^DOG(_)], which has only an occasional dyadic concept that has been "sealed in."

This passage usefully contrasts the conceptual structures assembled by FL with the those that are often assumed by linguists—wrongly, by Pietroski's lights—to be available to humans *antecedent* to the development of language.

2.4.3. *Concepts, predicative and sentential*

We are now in a position to ask more specific questions about Pietroski's third key notion—viz., that of a concept. As we've already seen, he takes these to be expressions in a compositional language of thought, some of which can be assembled by the semantic module of FL. But, however they might be assembled, they are the representations that allow us to think about the world.

> [C]oncepts have contents that can be described as ways of thinking about things; cf. Evans. A concept that can be used to think about something as a rabbit, whatever that amounts to, has a content that we can gesture at by talking about the concept type RABBIT. An instance of this type is a mental symbol that can be used to think about a rabbit as such, or to classify something—perhaps wrongly—as a rabbit; see Fodor. A concept of the type RABBIT-THAT-RAN, which can be used think about something as a rabbit that ran, is presumably a complex mental symbol whose constituents include an instance of RABBIT. A thought can be described as a sentential concept that lets us think about (some portion of) the universe as being a certain way. Thoughts of the type A-RABBIT-RAN can be used to think about the world as being such that a rabbit ran. (4)

As the remarks at the end of this passage indicate, Pietroski takes thoughts to be a *special kind of concept*—namely, a *sentential* concept. This is important to highlight, in view of its relation to a broader point about sentential *meanings*.

Pietroski is skeptical that "Slangs generate sentences as such." The traditional notion of a sentence, as a unity of a subject and a predicate, has been roundly abandoned in contemporary linguistics. While the notions of "subject" and "sentence" have a place in subject-predicate conceptions of *thought*, Pietroski points out that they "may have no stable place" in contemporary scientific grammars (114).

> Linguists have replaced "S" with many phrase-like projections of functional items that include tense and agreement morphemes, along with various complementizers. This raises questions about what sentences are, and whether any grammatical notion corresponds to the notion of a truth-evaluable thought. But theories of grammatical structure—and to that extent, theories of the expressions that Slangs generate—have been improved by *not* positing a special category of sentence. So while such a category often plays a special role in the stipulations regarding invented formal languages, grammatical structure may be independent of any notion of sentence. (61)

Accordingly, Pietroski suspects that talk of "grammatical subjects" is just a roundabout way of "saying that tensed clauses have a 'left edge constituent' that somehow makes them complete sentences—whatever that amounts to—as opposed to mere phrases like 'telephoned Bingley'" (87). Rather than clarifying the notion of a "complete sentence," Pietroski points out that talk of grammatical subjects *presupposes* it.

How, then, to characterize sentences? Naturally, Pietroski does *not* appeal to a distinction between sentential truth conditions and subsentential satisfaction conditions. Instead, he develops a novel version of predicativism, according to which *all* of the concepts assembled by *Begriffsplans* are predicative, in the sense that they all have a classificatory function. This includes concepts that are fetched by linguistic expressions like 'Jessica', 'David Pereplyotchik', and 'Reykjavík'. (Yes, *the* Reykjavík.)

So far, the view on the table is a version of the familiar predicativist position that was introduced by Quine (1970), defended by Burge (1973), and reanimated in contemporary discussions by the work of Delia Graff Fara (2005). Pietroski goes on, however, to make a quite novel claim—namely, that the concepts assembled by *sentence*-sized Slang expressions are *also* predicative.

The idea is that familiar subsentential predicates are assembled, largely via predicate conjunction, and then a new mental operation (⇑ or ⇓) converts the results into a sentential predicate—what Pietroski calls a "polarized concept". Here is how he defines these: "Given any concept M, applying the operation ⇑ yields a polarized concept, ⇑M, that applies to each thing if M applies to something" (30). For instance, if RABBIT applies to something, then ⇑RABBIT applies to each thing and ⇓RABBIT applies to no-thing. We will return to this topic in §3, when we compare this proposal with the inferentialist account of sentence meaning.

Recall that semantic instructions (*Begriffsplans*) have "mechanical execution conditions". Because Pietroski takes *Begriffsplans* to *be* linguistic meanings, it follows for him that that "meanings satisfy demanding compositionality constraints." Such constraints, he argues, permit the assembly of concepts that are better suited for their role in language use than for the epistemic role of "fitting the world". This important upshot of Pietroski's view bears on his rejection of both Davidson's extensional semantics and Lewis's unrestricted type-theoretic approach to natural language (§2.1). For, although he leaves it open that we might build truth-evaluable thoughts as a *side-effect* of language processing, he denies that "meanings are instructions for how to build concepts that exhibit classical semantic properties" (115). Likewise, he suspects that "most natural concepts [do not] have extensions; cp. Travis… if only because of vagueness; cp. Sainsbury" (9). Hence, the *Begriffsplans* that Pietroski identifies with meanings "make no reference to the things we usually think and talk about" (115). If correct, this conclusion is just one more nail in the coffin of the extensionalist project.

2.5. *Pietroski on Fregean thoughts and concepts*

Common to both Pietroski and Brandom is a deep engagement with the work of Frege. However, as we'll see presently, the lessons that Pietroski draws from Frege are *not* those that one might expect. In particular, the formal device that he takes over from Frege's semantics is not that of function application, as is common; rather, he emphasizes Frege's immensely useful notion of concept *invention*—something you don't hear much about in discussions of Frege, at least amongst linguists.

As noted earlier, Pietroski holds that are multiple languages of thought—i.e., distinct formats of concept application. In his discussions of Frege, he advances the hypothesis that there are, in fact, at least two such languages. The first one, in order of evolutionary history, may well have a Fregean semantics and include expressions of type <t>. The second one, which only came in with the evolution of natural language, consists of concepts that were *invented*, or *introduced*, in a Fregean sense, on the basis of the older ones.

> [N]atural sentences of type <t> may belong to languages of thought that are phylogenetically older than Slangs. Expressions of these newer languages may be used to build complex monadic concepts, perhaps including some special cases that are closely related to natural thoughts of type <t>. In which case, the very idea of a truth-conditional semantics for a human language may be fundamentally misguided. (114)

Because Pietroski treats the new type of concept as being invariably predicative—i.e., functioning semantically to classify things into *categories*, not to denote them individually—he calls such concepts "categorical". The older type of concept, which participates in thoughts of type <t>, includes singular denoting concepts and predicates of any adicity. On account of their semantic function of relating items to each other, Pietroski calls such concepts (and the thoughts they participate in) "relational".

Though I see its significance, I'm not, myself, a huge fan of the 'categorical'/'relational' terminology. Adverting to their historical roles, rather than their internal logic, I'll call these languages Olde Mentalese and New Mentalese for the remainder of the discussion. Here's how Pietroski casts the theoretical relations between them.

> Frege assumed that we naturally think and talk in a subject-predicate format, and that we need help—[e.g.] his invented *Begriffsschrift*—in order to use our rudimentary capacities for relational thought in systematic ways… The idea was that a thought content can be "dimly grasped," in some natural way, and then re-presented in a more logically perspicuous format that highlights inferential relations to other contents… I think this is basically right: our categorical thoughts are governed by a natural logic that lets us appreciate certain implication relations among predicates; but our relational concepts are related in less systematic ways. We use relational concepts in natural modes of thought. (95-6)

The distinction between Olde Mentalese and New Mentalese allows Pietroski clarify his perspective, contrasting it with Frege's. Here, too, it's instructive to quote at length.

> Frege introduced higher-order polyadic analogs of monadic concepts. In this respect, my project is the converse of his. Frege invented logically interesting concepts, and he viewed monadicity as a kind of relation to truth, as part of a project in logic that prescinds from many details of human psychology. I think humans naturally use concepts of various adicities to introduce logically boring predicative analogs. But I adopt Frege's idea that available concepts can be used to introduce formally new ones, and that this can be useful for certain derivational purposes. Frege "unpacked" monadic concepts like NUMBER(_), in ways that let him exploit the power of his sophisticated polyadic logic to derive arithmetic axioms from (HP). I am suggesting that Slangs let us use antecedently available concepts— many of which are polyadic—to introduce concepts like CHASE(_) and GIVE(_), which can be combined in simple ways that allow for simple inferences like conjunction reduction. But the big idea, which I am applying to the study of Slangs, is Fregean: languages are not mere tools for expressing available concepts; they can be used to introduce formally new concepts that are useful given certain computational capacities and limitations. This is why I have dwelt so long on Frege's project. For while the idea of concept introduction was important for Frege, it is not the aspect of his work that semanticists typically draw on.

The gory details of Frege's technical devices for concept introduction are, mercifully, beyond our present needs; only a few key points are relevant. One is that introducing concepts need *not* be seen on the model of explicit definition. Rather, Pietroski highlights Frege's proposal for a *second* way of introducing concepts—viz., by *inventing* them. Similarly, although *analyzing* a concept has often been seen as breaking it down into its more basic definitional constituents, Pietroski joins Fodor (1970) in rejecting the idea that lexicalized concepts will generally admit of such analytic definitions. Nevertheless, there is an alternative way of analyzing concepts, which Pietroski characterizes as "a *creative* activity" (emphasis mine).

> Given a very fine-grained notion of content, or thought-equivalence, analysis may not be possible. But Frege employed at least two notions of content: one based on his notion of sense (*Sinn*), and another according to which thoughts are equivalent if each follows from the other. Given the latter notion, or Lewis's characterization of contents as sets of logically possible worlds, one can say that our current representations are not yet perspicuous. We can use our concepts to ask questions that lead us to reformulate the questions in ways that allow for interesting answers. From this perspective, analysis can be a creative activity whose aim is not to depict our current representations…

It's in virtue of our ability to invent new concepts that we, *qua* humans endowed with a specific FL, have invented the monadic and quasi-dyadic concepts that arise only for language use. This includes not only monadic event-predicates like GIVE(_), invented on the basis of the older triadic concept GIVE(*x, y, z*), but also—importantly for Pietroski's pur-

poses, though not ours—*thematic* concepts such as AGENT(_), PATIENT(_), and RECIPIENT(_).

2.6. *Summary*

The generativist methodology that animates Pietroski's inquiry leads him to a number of strikingly original claims about concepts and a detailed theory of meanings. Treating the latter in a resolutely naturalist fashion, he maintains that their theoretical role is to mediate between pronunciations and concepts—i.e., to effect the psychological operations that constitute the interface between language (FL) and the "conceptual-intentional system" (to use Chomsky's coinage). Although meanings facilitate the *assembly* of concepts, which *have* intentional contents, Pietroski holds that meanings are neither concepts *nor* their contents.

On this view, the relation between truth and conceptual/intentional content is "quite complicated and orthogonal to the central issues concerning how meanings compose" (115). This, among the many other reasons surveyed above, leads Pietroski to abandon Davidson's project of extensional truth-conditional semantics. Moreover, the goal of *explaining* our access to a productive hierarchy of concepts, rather than merely *stipulating* it, underlies his rejection of the type-theoretic approach championed by Lewis (1970)—one of the many disagreements that we'll look at in the next section.

The semantic theory that satisfies Pietroski's methodological commitments—as well as the compositionality constraints that he argues follow from it—treats meanings as composable instructions for concept assembly. The instructions are "composable" in the sense that their basic constituents—namely, *fetch@* and *join[I, I']*—can enter into part-whole relations to one another. Moreover, as noted earlier, the larger structures they compose will, in a definite sense, *mirror* those of the concepts that the instructions assemble.

Having furnished empirical evidence for the idea that these "*Begriffsplans*" reduce largely to two flavors of *predicate conjunction*, Pietroski adopts a strong version of predicativism, according to which *all* of the concepts that natural language allows us to access and assemble are predicative. This includes not only the concepts fetched by linguistic expressions that have *traditionally* been classed as predicates, but also those that have generally been seen as differing in some important respect—including singular terms and, more strikingly, even *sentences*. The conceptual predicates that meanings allow us to access and assemble thus all either monadic, dyadic (in a restricted sense), or "polarized", where the latter kind is assembled by sentence-like linguistic expressions, using specialized mental operations, ⇑ and ⇓, to "polarize" concepts. Importantly, the resulting conceptual structures are not necessarily ones that best "fit the world", and they're not even the only ones we can deploy in thought. But, if Pietroski is correct, the they *are* the only ones that FL can assemble.

Denying that the concepts involved in language use have denotational properties and relational structures (of arbitrary adicity) leaves open whether *other* concepts might have these features. As we saw, Pietroski hypothesizes that there *are* in fact such concepts, and that they belong to a phylogenetically older language of thought than the one FL allows us to access—what I've dubbed 'Olde Mentalese'. Olde thoughts might have a subject-predicate form, a Fregean semantics, and belong to the semantic type <t>.

Pietroski goes on to make novel use of Frege's notion of concept *invention* in explaining the (non-definitional) mental introduction of *new* concepts on the basis of the Olde ones—specifically, the ones that FL allows us to access/assemble (New Mentalese). This psychological process, he argues, serves to introduce GIVE(_) on the basis of GIVE(*x, y, z*), as well as novel *thematic* concepts such as AGENT(_), PATIENT(_). These, in turn, participate in building polarized sentential concepts, such as ⇑RABBIT, which "applies to each thing if RABBIT applies to something". In the course of assembling such concepts, it may happen—*but only as a side-effect* (fortuitous or otherwise)—that we *also* token thoughts of Olde Mentalese. But the details of how Olde Mentalese thoughts functions are, Pietroski rightly holds, beyond the scope of a naturalistic semantic inquiry into human *language*.

Conclusion of Part 1

We've now surveyed the core commitments of two large-scale theoretical frameworks in the philosophy of language and seen some of the ways in which they play out in the realm of semantics, including in detailed analyses of various linguistic constructions. It may appear that the two views are so different in substance and overall methodology that a conversation between the two is unlikely to bear much fruit. In fact, I suspect this is a large part of why so few conversations of this kind ever take place. In Part 2 of this essay (next issue), I'll argue for a contrary perspective, outlining an ecumenical approach that seeks to integrate the two in a variety of ways. In surveying what I take to be significant points of convergence—which then serve as background for constraining residual disputes—I rebuff various superficial objections to the possibility of integration. In each case, I show how the theoretical differences that they point to can be reconciled without doing much (if any) violence to either view.

References

Brandom, R. B. 1987. "Singular Terms and Sentential Sign Designs." *Philosophical Topics* 15 (1): 145–167.

Brandom, R. B. 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge: Harvard University Press.

Brandom, R. B. 2001. *Articulating Reasons: An Introduction to Inferentialism*. Cambridge: Harvard University Press.

Brandom, R. B. 2008. *Between Saying and Doing: Towards an Analytic Pragmatism*. Cambridge: Harvard University Press.

Brandom, R. B. 2009. "Why Philosophy Has Failed Cognitive Science." In *Reason and Philosophy: Animating Ideas*. Cambridge: Harvard University Press.

Burge, T. 1973. "Reference and proper names." *Journal of Philosophy* 70: 425–439.

Burge, T. 2010. *Origins of Objectivity*. New York: Oxford University Press

Churchland, P. M. 1979. *Scientific Realism and the Plasticity of Mind*. Cambridge: Cambridge University Press.

Chomsky, N. 1986. *Knowledge of Language*. New York: Praeger.

Chomsky, N. 1995. *The Minimalist Program*. Cambridge: MIT Press.

Chomsky, N. 2000. *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Chomsky, N. 1995. *Why Only Us: Language and Evolution*. Cambridge: MIT Press.

Collins, J. 2012. *The Unity of Linguistic Meaning*. Oxford: Oxford University Press.

Davidson, D. 1983. *Essays on Truth and Interpretation*. Oxford: Oxford University Press.

Dummett, M. 1978. *Truth and other Enigmas*. London: Duckworth.

Fodor, J. A. 1975. *The Language of Thought*. Oxford: Oxford University Press.

Fodor, J. A. 1979. "Three Reasons for not Deriving 'Kill' from 'Cause to Die'." *Linguistic Inquiry* 1: 429-438.

Fodor, J. A., Bever, T. and Garrett, M. 1974. *The Psychology of Language*, Cambridge: MIT Press.

Fodor, J. A. and Lepore, E. 1992. *Holism: A Shopper's Guide*. Oxford: Blackwell.

Fodor, J. A. and Pylyshyn, Z. W. 2015. *Minds Without Meanings: An Essay on the Content of Concepts*. Cambridge: MIT Press.

Grover, D. L., Camp, J. L., Belnap, N. D. 1975. "A Prosententialist Theory of Truth." *Philosophical Studies* 27 (2): 73–125.

Graff Fara, D. 2015. "Names are Predicates." *Philosophical Review* 124 (1): 59–117.

Grice, H. P. 1989. *Studies in the Way of Words*. Cambridge: Harvard University Press.

Harris, D. W. 2020. "We Talk to People, not Contexts" *Philosophical Studies* 177 (9): 2713–2733.

Heim, I. and Kratzer, A. 1998. *Semantics in Generative Grammar*. Oxford: Blackwell.

Horwich, P. 1990. *Truth*. Oxford: Oxford University Press.

Kripke, S. 1980. *Naming and Necessity*. Cambridge: Harvard University Press.

Levinson, S. C. 1983. *Pragmatics*. Cambridge: Cambridge University Press

Lewis, D. K. 1969. *Convention*. Cambridge: Harvard University Press.

Lewis, D. K. 1970. "General Semantics." *Synthese* 22: 18–67.

Lewis, D. K. 1973. "Languages and Language." In K. Gunderson (ed.). *Minnesota Studies in the Philosophy of Science*. Vol. 7. Minneapolis: University of Minnesota Press.

Pereplyotchik, D. 2017. *Psychosyntax: The Nature of Grammar and Its Place in the Mind*. Springer International.

Pereplyotchik, D. 2019. "Whither Extensions." *Mind and Language* (Special Issue) 1–14.

Pietroski, P. M. 2018. *Conjoining Meanings: Semantics Without Truth Values*. Oxford: Oxford University Press.

Price, H. 2011. *Naturalism Without Mirrors*. Oxford: Oxford University Press.

Price, H. 2013. *Expressivism, Pragmatism and Representationalism*. Cambridge: Cambridge University Press.

Quine, W. V. 1953. "On What There Is." In *From a Logical Point of View*. New York: Harper and Row.

Quine, W. V. 1960. *Word and Object*. Cambridge: MIT Press.

Quine, W. V. 1970/1980. *Philosophy of Logic, 2nd Edition.* Cambridge: Harvard University Press.

Rorty, R. 1979. *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.

Sellars, W. 1963. "Philosophy and the Scientific Image of Man." In R. Colodny (ed.). *Science, Perception, and Reality*. Humanities Press/Ridgeview: 35–78.

Stainton, R. J. 2006. *Words and Thoughts: Subsentences, Ellipsis, and the Philosophy of Language*. Oxford: Oxford University Press

Stalnaker, R. 1984. *Inquiry*. Cambridge: MIT Press.

Sperber, D. and Wilson, D. 1986/1995. *Relevance*: *Communication and Cognition*, *2nd Edition*. Oxford: Blackwell.

Tomasello, M. 2005. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge: Harvard University Press.

Yang, C. 2006. *The Infinite Gift: How Children Learn and Unlearn the Languages of the World*. New York: Scribner.

Croatian Journal of Philosophy
Vol. XX, No. 60, 2020

But Without …? Reflections on Pietroski's Conjoining Meanings

MICHAEL GLANZBERG*
Rutgers University, New Brunswick, New Jersey, USA

In this short note, I discuss the viability of truth-conditional semantics in light of Pietroski's criticisms. I explore an alternative view that follows Pietroski in putting emphasis on the relation of meanings to concepts, but makes some room for truth conditions.

Keywords: Lexical meaning, quantifiers, concepts, minimalist program, language faculty.

Paul Pietroski's book *Conjoining Meanings* (Pietroski 2018) is the culmination of a long research project.[1] The goals of the project are ambitious. One is to offer a forceful and thorough challenge to the program of truth-conditional semantics; indeed, Pietroski's goal is to convince us to give up on this project. Semantics is a broad field, and maybe it is fool-hardy to try to say what the dominant approach to semantics is, but in the areas of formal linguistics and philosophy of language, the truth-conditional program is central, mainstream, and perhaps seen as the only game in town. Pietroski challenges that program from the very perspective of formal linguistics and philosophy of language, so his challenge is all the more powerful. But in addition, Pietroski offers us an alternative approach, that does away not only with truth conditions, but a great deal of the formal apparatus that typically goes with truth-conditional semantics; especially, does away with type theory.

* Thanks to Ernie Lepore and Paul Pietroski for extended discussions of Paul's work during a reading group we held in 2019. Thanks also to John Collins for many discussions of ideas closely related to Paul's.

[1] Building on such work as Pietroski (2003, 2005, 2010, 2012).

One of the important features of Pietroski's positive proposal is a link between meanings and concepts. Pietroski's proposal thus offers a certain rather specific kind of internalism about meaning, in line with a broadly Chomskian view of the language faculty and the place of semantics in it.

For some time now, I have been working on a project that was always inspired by Pietroski's work; and especially, agrees with the general idea that meanings involve relations between words and concepts. But unlike Pietroski, I have offered this project as a way of supporting a limited form of truth-conditional semantics, and showing how truth-conditional semantics can interact fruitfully with the broader cognitive sciences.[2]

You might imagine two builders assessing an old and much-loved building that is showing signs of falling over. I find myself suggesting that if we shore up the foundations and do some structural repairs, we can keep an old lovely building in its glory. Pietroski, we might say, suggests that the case is hopeless, and the responsible thing to do is to knock it down and put up something better in its place.

To continue our metaphor. The strange thing is I have repeatedly found, when Pietroski and I make our proposals to the client, that I get asked what the difference in result really is. In this short note, I shall try to illustrate a few of the key differences. I shall not really try to argue which is right, just explain where there are choices.

1. *Agreeing and agreeing to disagree*

There is one point where Pietroski and I agree fully. We have both emphasized the importance of *concepts* for lexical meaning. The idea here is simple and familiar. A child learning their language must learn to associate the sound / dɔg / with a meaning. It is a common assumption in a great deal of work on language acquisition that they do so by associating that sound with a concept, DOG. Thus, the main thing that gives the meaning of *dog* is the concept DOG.

As a kind of motivating idea, there is much to like about this proposal. But, there are many many reasons to be dissatisfied with it in detail.[3] Depending on what one means by 'concept' this might be far too narrow an idea to account for the range of word meanings. And of course, the implicit supposition in the story is that the learner has the concept already in place, and then associates a sound with it. That may be true in some cases, but it is dubious as an explanation of all our word leaning.

All of those are points where Pietroski and I are happy to agree, both about what is easy and hard about the main idea. There is, how-

[2] For instance, Glanzberg (2011, 2014, 2018).

[3] A nice illustration is a handbook article by Clark (1983), who uses the idea as an introduction, but almost immediately takes it back.

ever, one point of likely disagreement over this leading idea. Pietroski is friendly to a kind of Fodorian atomism about concepts (Fodor 1975, 1998), while I have stressed the importance of the internal structure of concepts for how concepts relate to truth-conditional semantics (Glanzberg 2018). For this discussion, I shall simply put aside this difference.[4]

Another point of disagreement is much more central, but I shall also put it aside for now. A great deal of Pietroski's view is motivated by issues of polysemy. Because of this, he most definitely does not say that we link a sound to a single concept. Rather, we link a sound to an address, at which a cluster of concepts is stored. Any one of them may be selected. Hence, we get for most every word, a family of polysemous meanings. This is important, as many of Pietroski's arguments against truth-conditional semantics stem from problems of polysemy. I shall set this issue aside too. Not because I think it is a small matter, but because I think there are a range of issues to focus on once we resolve polysemy.

There are also, of course, questions about how much internalism about meaning is correct. Both Pietroski and I are opting for something more internalist than many philosophers would expect (though many cognitive scientists would find entirely obvious). I think there are interesting questions about how far the internalist motivations about meaning go, but again, I shall not pursue them here.

There is another family of issues that I think is well worth pursuing, and is important to Pietroski's work, but I shall not explore in any depth here. This is the family of issues about what the right tools are for semantics. In particular, as Pietroski rightly notes, a great deal of truth-conditional semantics, perhaps since Montague (1973) and certainly since Partee (1975), is done using the apparatus of the simple theory of types, and the treatment of variables in semantics is broadly Tarskian, in the tradition of Tarski (1935).[5] Pietroski gives an extended discussion of why he finds this to be a fundamental mistake.

Here, I do disagree, but again, I shall not say much about why. To go back to our metaphor of the builders, this is an important issue between builders, but not of much interest to the client. Which tools work, which are dangerous, and what they produce is important to builders, but it is only of interest to the client to the extent that they produce visibly different results.

Another point along these lines is about the technical issue of lexical decomposition. I have described my view of the lexicon as a 'pointers and packaging' approach, which has each lexical entry point to a concept, but also contain a great deal of specifically grammatical structure that 'packages' the concept into a word meaning (Glanzberg 2018). For illustration's sake, I have presented this in the form of a lexical de-

[4] For an overview of work on concepts, see Murphy (2002).

[5] Some programs make even strong use of type theory. See, for instance, van Benthem (1991).

composition, following the tradition of Dowty (1979) and more recently Rappaport Hovav and Levin (1998). Pietroski again disagrees, but in a specific technical way. He does not have much (if any) packaging in the lexicon. But, he has a great deal of syntax that builds complex 'words' from simpler roots, following a different tradition, from Borer (2005), Hale and Keyser (2002), and Ramchand (2008). So, he does not really reject all packaging, but does put it in a different place. There are lots of questions here, both empirical and methodological. But again, they are more a matter between builders than for the client.

What I think is of interest to the client is what the result is. In this case, if we think that meanings link to concepts in the right way, and that concepts are at first pass internal mental representations, what is left for truth-conditional semantics?

2. *Go fetch?*

Pietroski's core idea is that meanings are instructions to, as he puts it, *fetch* a concept from a given address in long-term memory. As I noted, he argues that at any such address is a family of concepts, but still the main semantic instruction is, as he puts it ***fetch***. To fill in a little more, each lexical item gives an instruction to fetch a concept at an address, so the meaning of *cow* is the instruction ***fetch***@cow. Find the address linked to *cow*, and fetch a concept from there.

The issue I want to focus on is what fetches. For Pietroski, that is the core semantic operation. So, any meaningful morpheme will fetch. I shall suggest another way to look at this.

As is well-known, 'words'—morphemes, lexemes, or whatever our theory tell us to use—come in two classes. There are open and closed classes, and maybe a few in-between. Open classes are just they sound like: open. We can add to them as our hearts and interests desire. In English, and many languages, these correspond to the major lexical categories: nouns, verbs, and adjectives/adverbs. We can add nouns and verbs easily. Examples are familiar. *Carburator, transistor, transduce* are relatively late additions to English. Looking at the *Oxford English Dictionary's* new word list for 2020 I found *'hend', adv. and prep., sense 2: "In a diligent or skillful manner; adeptly, nimbly"*.[6] (Apparently some link to older phrases for knighthood.) The open classes of words grow, and they seem to grow along with our concepts. Discovering the concepts of transduction or mitosis help us to make the words *transducer* and *mitosis*.

But there are also closed-class words. Expressions that give us tenses, quantifiers, moods, and a number of other 'grammatical' terms form stable classes, that do not change; or if they do, they change at the glacial pace of grammatical change. We cannot add a new tense or quantifier to our languages the way we can add a new noun or verb.

[6] Accessed at https://public.oed.com/updates/new-words-list-january-2020/.

Perhaps the more natural class here is what linguists call *functional* items. These are roughly grammatical elements, like tenses, moods, quantifiers, but also number markers, light verbs, and many more. They overlap with the closed classes more or less. (Whether pronouns or prepositions are functional can be argued. That is perhaps the main potential difference.)

There are two hypotheses we can entertain about the functional expressions, and this marks an important point of difference between my view and Pietroski's. One option is that functional elements fetch concepts, like most other morphemes. I think this is Pietroski's view. Now, these need not be entirely ordinary fetches. The addresses at which these expressions fetch may have a more limited or more specialized range of concepts. So, we may find less polysemy. The concepts involved may have different sources than other concepts, and they may be special in other ways. Pietroski explores these questions in depth. But in the end, the semantic job of a functional expression is to fetch.

There is another option, and it is one I have endorsed. Functional elements do not fetch. In effect, their meanings are part of the grammar, and we not need to fetch anything extra-grammatical to provide them.

To make this clear, let me say a little more about another very high-level assumption that is common across Pietroski and me. We both adopt a broadly Chomskian view of language. This is far too big an issue to state quickly. But the main idea is clear enough. There is a language faculty. This is a part of human cognition. It is a distinctive cognitive system. It is one that is substantially innate. Most importantly, it is one whose principles and parameters make up Universal Grammar (Chomsky 1986, 2000; Collins 2004).

So, with this background, we can ask a reasonably clear question about what is part of grammar: what is encoded in the language faculty. It should be clear enough that the full force of Chomsky's views is not really required to ask this question. So, for instance, just what is innate is not immediately at stake. Any reasonably strong domain-specificity for grammar will suffice, but Chomsky's views are common ground between Pietroski and me, so we might as well go for the stronger hypothesis.[7]

With this idea in mind, we can think about ***fetch***-ing. As I understand it, this operation asks an expression to link to something outside of the language faculty proper. This is not surprising with lexical items. The source of words like *carburator* is not our native linguistic ability, it is our extra-linguistic ability to build, think about, and then talk about, cars and their parts. So, the instruction to ***fetch***@carburator is a link to something outside the language faculty.

As I have said, I think this is plausible, with some minor disagreements, for the major lexical categories. But should we extend it to the

[7] For some thoughts on domain-specificity, see among a huge literature Hirschfeld and Gelman (1994), and the many references therein.

functional ones? Here we reach the first point of disagreement I shall highlight. I, and many proponents of truth-conditional semantics within a Chomskian framework, would argue that the meanings of some functional elements are within the language faculty, and so are strictly part of grammar.

A second question is what the meanings of those sorts of expressions are like. Functional expressions are the ones that are discussed in most detail in standard expositions of truth-conditional semantics. It is thus a well-motivated hypothesis that the tools and methods of truth-conditional semantics are the right ones to describe their meanings. The two ideas combine, to indicate that substantial parts of truth-conditional semantics describe parts of grammar, by describing the meanings of the functional elements.

We thus seem to have a very stark difference in views. On the one hand, we have a language faculty filled with semantics in a truth-conditional style. Perhaps not as much as some views of semantics might suppose, but nonetheless, a rich semantics within the language faculty. On the other, we have a highly restricted language faculty, with a very few semantic operations, whose main job is to access extra-linguistic concepts, and combine them in simple ways.

But as we begin to look at what differences these two starkly different pictures make in practice, it becomes harder to distinguish them. The point is illustrated from both sides. From my pro-truth-conditions side, I freely admit that the notion of truth conditions is being stretched rather far. All I claimed is that the familiar apparatus of truth-conditional semantics, applied carefully to specific points, is useful. As is well-known, this apparatus is quite rich and flexible, and so one might well ask how substantial such a general claim is. At the same time, from Pietroski's side, we see a range of concepts that are closely tied to grammar, like closure operators and indices, and so on. Of these, Pietroski comments that we find "expressions that can be used to build T-concepts, which bear an intimate relation to certain truth-evaluable *thoughts*" (316). And in a number of cases the glosses on those concepts are close in nature to what the truth-conditional program would say. For instance, Pietroski (with due caution) endorses a Reichenbachian account of tense. This is one among several that can be articulated with standard truth-conditional apparatus. So, in the abstract we might clearly distinguish between a semantically rich language faculty or a semantically minimal one. But in practice, we see something much harder to identify: it might be a rich language faculty which makes special and partial use of truth-conditional apparatus, or a semantically sparse language faculty that creates strong links to extra-linguistic concepts that have exactly the same truth-conditional properties.

Let us look at one specific case: quantifiers. This is a good case for my side, as most any textbook on truth-conditional semantics will have

a great deal to say about quantifiers.[8] Even more, Partee (2015) marks discovering the importance of quantifiers as a major event in the development of semantics in generative grammar. I myself have used quantifiers as an example of where we get strong truth-conditional results (Glanzberg 2014).

The classic theory of determiner meanings as generalized quantifiers from the early 1980s (Barwise and Cooper 1981; Keenan and Stavi 1986; Higginbotham and May 1981) was indeed an impressive achievement. It offers us meanings for the interesting closed class of determiners (in languages like Germanic ones that have lots of determiners). We get meanings for English *all, some, most, few, ….* And of course, we get them couched in the mechanisms of truth-conditional semantics.

Here again, there are a number of questions we should pause to ask. Though generalized quantifier theory was an important step, a great deal of more recent work, both empirically and theoretically oriented, has shown its limits and weaknesses. It is by no means that last word on the semantics of quantifiers.[9] And again, Pietroski does not disagree on the basic meanings of quantifier expressions, nor on the need for some grammar to go with quantifiers (movement, indices, etc.). So, again, where does the larger disagreement show up in practice?

There are a number of more theory-specific points where disagreement becomes sharper. One is about what machinery to use to give the meanings of quantifiers. Standard generalized quantifier theory is embedded in the simple theory of types, which Pietroski finds unreasonably powerful. He observes that the familiar meanings for many generalized quantifiers can also be given in monadic second-order logic. He prefers the Boolos-inspired plural interpretation of second-order logic. I, in contrast, worry that second-order logic itself is far too powerful, even under the plural interpretation. As is well know, second-order logic has a sentence of pure logic that is a logical truth just in case the continuum hypothesis of set theory is true, for instance (see Shapiro 1991). So here, at least within two research programs, we have a genuine difference in what tools to use.[10] There are also some interesting questions about how to explain some important facts about quantifier meaning, such as the well-known conservativity constraint. A common idea is that this is a semantic universal, and so is simply 'hard-wired' into the language faculty (cf. von Fintel and Matthewson 2008). This is not really much of an explanation, of course; rather, a claim that no further explanation will be found. Pietroski is not satisfied with this view, and makes a (tentative) suggestion about how a better explanation might be found.

[8] Heim and Kratzer (1998) is a much-cited example, but most semantics textbooks do the same.

[9] See Beghelli and Stowell (1997), Landman (2004), Reinhart (1997), Szabolcsi (2010), and Wellwood (2019), among many others.

[10] See Boolos (1984), and for subsequent discussion, see Shapiro (1991) and critical discussion by Jané (2009).

But what we find here is not disagreement on the basic meanings of quantifiers, and between Pietroski and me, not really much disagreement about the grammar either. Again, we find two sorts of disagreement. One is about tools: where they work, how to use them. This is very much a theory-internal kind of disagreement. To return to the builders metaphor yet again, it is a disagreement that is mostly between builders. Our client may not really care. The other is about big-picture issues about what is part of grammar proper. But again, we may find that our client does not really see the difference in practice, and may not worry about it quite to the extent that we do.

3. *Get with the program?*

My brief and casual discussion of a few differences between Pietroski's view and my own reveal two sorts of differences. On the detailed, theory-internal side, there are lots of questions about which methods, tools, and analyses are correct. Here, Pietroski argues in favor of a radical departure from the truth-conditional program, while I argue for judicious modification to keep the truth-conditional approach. In her elegant review of Pietroski (2018), Ramchand (2020) suggests many of us will respond to Pietroski's taking our beloved truth-conditional semantics from us by going through stages of grief. She identifies a bargaining stage, and it would appear I am offering to do just that kind of bargaining. I shall, as she puts it, use the traditional tools of truth-conditional semantics, but where and how they work best. I indeed am bargaining, but I think the bargain is a good one.

I suppose I also hold out some more optimism for the prognosis of the patient. Where Pietroski (and Ramchand) are sure the situation is grave, I keep hoping for a turn-around. So, I am more optimistic than Pietroski about how the internal structure of concepts can yield extensions, and more optimistic about fitting a restricted range of empirically robust composition principles within a type-theoretic framework. It is easier to bargain when you feel optimistic about the outcomes.

The other major point of disagreement is more abstract; perhaps more philosophical than methodological or empirical. As I mentioned, my preferred view ends up with a language faculty that is rather rich in semantics, and includes a great deal of truth-conditional apparatus. Pietroski ends up with a very different result. His language faculty has little semantics beyond ***fetch***. Where I see semantically rich elements of grammar, Pietroski sees elements of grammar that ***fetch*** in specific ways.

I think it is helpful to frame this disagreement in terms of some recent thinking about syntax, and more widely, grammar. I have in mind the minimalist program, following, among many authors, Chomsky (1995, 2000), or Hornstein (1995). The minimalist program, as a research program, has many components. Some are developments in syntax that have received broad acceptance across a range of approaches to syntax and semantics. Others are more ambitious, and more controversial.

Pietroski endorses the general idea of keeping the language faculty as simple as possible; in particular, to posit as limited a range of operations within the language faculty as possible. He is careful not to endorse any particular version of the minimalist program, though he is clearly sympathetic. He writes "I do find "minimalist" conceptions of syntax attractive on empirical and conceptual grounds" (295).

But, within work in the minimalist tradition, we can find a very stark view of what goes into the language faculty, and such a view makes the difference between Pietroski's position and my own equally stark. An example can be found in Hauser, Chomsky, and Fitch (2002). They offer a proposal about what they call the *faculty of human language—narrow sense*, which is extremely minimal. According to their view, it is little more than an engine that supports recursion, with only whatever combinatorial apparatus is needed to enable recursion. Presumably, that includes something like *merge* as described by other minimalist work, and not much more. It is easy to see that the kinds of semantic mechanisms I have suggested belong to the language faculty are highly unlikely to be part of this sparse faculty. It is much more likely that the kinds of mechanisms that Pietroski proposes could be part of it.

Hauser, Chomsky, and Fitch (2002) also discuss what they call the *faculty of human language—broad sense* which includes some central interface systems, including interfaces with conceptual-intentional and sensory-motor systems. It is not an accident that a fair bit of what traditionally falls within the scope of linguistics falls within the broad, but not the narrow, faculty. The narrow faculty is an extremely minimal recursion engine.

The faculty of human language—narrow sense gives us a very stark picture of what is core to human language. The more we think that is central to grammar, the more unlikely the semantically rich language faculty I have advocated becomes. I think an often unspoken assumption of a great deal of work in truth-conditional semantics in the tradition of generative grammar is that the language faculty, in whatever sense is relevant, is broader than the very stark version offered by faculty of human language–narrow sense from Hauser, Chomsky, and Fitch.[11] Though fully deciding what goes into a narrow or broad language faculty is no easy task, Pietroski's option is much more likely to fit with a strong minimalist view.

Of course, knowing what really goes into the language faculty (narrow or broad) as opposed to related aspects of cognition, is no easy task. Hence, I think the difficulty in finding clear markers of the practical difference between Pietroski's proposal and mine is not so surprising. If we knew better how to probe for what is in the language faculty, perhaps clearer answers would be forthcoming.[12]

[11] Sometimes this is clearly articulated. See, for instance Larson and Segal (1995) and Ludlow (2011).

[12] But, for some thoughts about this, see Crain, Gualmini, and Pietroski (2005) and Pietroski and Crain (2012).

Absent that, I suppose we are speculating, based on what looks like good data and successful theories. My speculations go in a rather less minimal direction, while Pietroski's go more minimal. Mine go more optimistic about the value of familiar truth-conditional apparatus, his go rather more pessimistic. There are some clearer disagreements about tools and some specific data, but they are highly specific, and somewhat project-internal disagreements. To end again with the builders metaphor, we could forgive our client from having trouble seeing just what for them the difference comes to. Absent a sharper understanding of the language faculty, that difference remains elusive.

References

Barwise, J. and Cooper, R. 1981. "Generalized Quantifiers and Natural Language." *Linguistics and Philosophy* 4: 159–219.

Beghelli, F. and Stowell, T. 1997. "Distributivity and Negation: The Syntax of *Each* and *Every*." In A. Szabolcsi (ed.). *Ways of Scope Taking*. Dordrecht: Kluwer: 71–107.

van Benthem, J. 1991. *Language in Action*. Amsterdam: North-Holland.

Boolos, G. 1984. "To Be Is to Be a Value of a Variable (or to Be Some Values of Some Variables)." *Journal of Philosophy* 81: 430–99.

Borer, H. 2005. *Structuring Sense*. Oxford: Oxford University Press.

Chomsky, N. 1986. *Knowledge of Language*. New York: Praeger.

———. 1995. *The Minimalist Program*. Cambridge: MIT Press.

———. 2000. *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Clark, E. V. 1983. "Meaning and Concepts." In P. H. Mussen (ed.). *Handbook of Child Psychology*. Fourth Edition, Volume III: Cognitive Development. New York: Wiley: 787–840.

Collins, J. 2004. "Faculty Disputes." *Mind and Language* 19: 503–33.

Crain, S., Gualmini, A. and Pietroski, P. M. 2005. "Brass Tacks in Linguistic Theory: Innate Grammatical Principles." In S. Laurence, P. Carruthers, and S. Stich (eds.). *The Innate Mind: Structure and Contents*. Oxford: Oxford University Press: 175–97.

Dowty, D. R. 1979. *Word Meaning and Montague Grammar*. Dordrecht: Reidel.

von Fintel, K., and Matthewson, L. 2008. "Universals in Semantics." *Linguistics Review* 25: 139–201.

Fodor, J. A. 1975. *The Language of Thought*. New York: Crowell.

———. 1998. *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.

Glanzberg, M. 2011. "Meaning, Concepts, and the Lexicon." *Croatian Journal of Philosophy* 31: 1–29.

———. 2014. "Explanation and Partiality in Semantic Theory." In A. Burgess and B. Sherman (ed.). *Metasemantics: New Essays on the Foundations of Meaning*. Oxford: Oxford University Press: 259–92.

———. 2018. "Lexical Meaning, Concepts, and the Metasemantics of Predicates." In D. Ball and B. Rabern (eds.). *The Science of Meaning: Essays on the Metatheory of Natural Language Semantics*. Oxford: Oxford University Press: 97–225.

Hale, K. and Keyser, S. J. 2002. *Prolegomenon to a Theory of Argument Structure*. Cambridge: MIT Press.

Hauser, M., Chomsky, N. and Fitch, W.T. 2002. "The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?" *Science* 298: 1569–79.

Heim, I. and Kratzer, A. 1998. *Semantics in Generative Grammar*. Oxford: Blackwell.

Higginbotham, J. and May, R. 1981. "Questions, Quantifiers and Crossing." *Linguistics Review* 1: 41–79.

Hirschfeld, L. A. and Gelman, S. A. (eds.). 1994. *Mapping the Mind: Domain Specificity in Cognition and Culture*. Cambridge: Cambridge University Press.

Hornstein, N. 1995. *Logical Form: From GB to Minimalism*. Oxford: Blackwell.

Jané, I. 2009. "Higher-Order Logic Reconsidered." In S. Shapiro (ed.). *Oxford Handbook of Philosophy of Mathematics and Logic*. Oxford: Oxford University Press: 781–810.

Keenan, E. L. and Stavi, J. 1986. "A Semantic Characterization of Natural Language Determiners." *Linguistics and Philosophy* 9: 253–326.

Landman, F. 2004. *Indefinites and the Type of Sets*. Oxford: Blackwell.

Larson, R. K. and Segal, G. 1995. *Knowledge of Meaning*. Cambridge: MIT Press.

Ludlow, P. 2011. *The Philosophy of Generative Linguistics*. Oxford: Oxford University Press.

Montague, R. 1973. "The Proper Treatment of Quantification in Ordinary English." In J. Hintikka, J. Moravcsik, and P. Suppes (eds.). *Approaches to Natural Language*. Dordrecht: Reidel: 221–42. Reprinted in R. H. Thomason (ed.). *Formal Philosophy*. New Haven: Yale University Press, 1974.

Murphy, G. L. 2002. *The Big Book of Concepts*. Cambridge: MIT Press.

Partee, B. H. 1975. "Montague Grammar and Transformational Grammar." *Linguistic Inquiry* 6: 203–300.

———. 2015. "The Garden of Eden Period for Deep Structure and Semantics." In Á. Gallego and D. Ott (eds.). *50 Years Later: Reflections on Chomsky's Aspects*. Cambridge, MA: MITWPL: 187–98.

Pietroski, P. M. 2003. "The Character of Natural Language Semantics." In A. Barber (ed.). *Epistemology of Language*. 217–56. Oxford: Oxford University Press.

———. 2005. *Events and Semantic Architecture*. Oxford: Oxford University Press.

———. 2010. "Concepts, Meanings and Truth: First Nature, Second Nature and Hard Work." *Mind and Language* 25: 247–78.

———. 2012. "Semantic Monadicity with Conceptual Polyadicity." In M. Werning, W. Hinzen, and E. Machery (eds.). *Oxford Handbook of Compositionality*. 129–48. Oxford: Oxford University Press.

———. 2018. *Conjoining Meanings: Semantics Without Truth Values*. Oxford: Oxford University Press.

Pietroski, P. M. and Crain S. 2012. "The Language Faculty." In E. Margolis, R. Samuels, and S. P. Stich (eds.). *Oxford Handbook of Philosophy of Cognitive Science*. Oxford: Oxford University Press: 361–81.

Ramchand, G. 2008. *Verb Meaning and the Lexicon*. Cambridge: Cambridge University Press.

———. 2020. "Truth Is Dead; Long Live the Truth. Commentary on *Conjoining Meanings* by Paul Pietroski." *Mind and Langauge* 35: 251–65.

Rappaport Hovav, M. and Levin, B. 1998. "Building Verb Meanings." In M. Butt and W. Geuder (eds.). *The Projection of Arguments*. Stanford: CSLI Publications: 97–134.

Reinhart, T. 1997. "Quantifier Scope: How Labor Is Divided Between QR and Choice Functions." *Linguistics and Philosophy* 20: 335–97.

Shapiro, S. 1991. *Foundations Without Foundationalism: A Case for Second-Order Logic*. Oxford: Oxford University Press.

Szabolcsi, A. 2010. *Quantification*. Cambridge: Cambridge University Press.

Tarski, A. 1935. "Der Wahrheitsbegriff in Den Formalizierten Sprachen." *Studia Philosophica* 1: 261–405. References are to the translation by J. H. Woodger as "The Concept of Truth in Formalized Languages" in Tarski, Alfred. 1983. *Logic, Semantics, Metamathematics*, Second Edition. Indianapolis: Hackett. Edited by J. Corcoran.

Wellwood, A. 2019. *The Meaning of More*. Oxford: Oxford University Press.

Croatian Journal of Philosophy
Vol. XX, No. 60, 2020

What Do We Experience When Listening to a Familiar Language?

ANNA DROŻDŻOWICZ*
University of Oslo, Oslo, Norway

What do we systematically experience when hearing an utterance in a familiar language? A popular and intuitive answer has it that we experience understanding an utterance or what the speaker said or communicated by uttering a sentence. Understanding a meaning conveyed by the speaker is an important element of linguistic communication that might be experienced in such cases. However, in this paper I argue that two other elements that typically accompany the production of spoken linguistic utterances should be carefully considered when we address the question of what is systematically experienced when we listen to utterances in a familiar language. First, when we listen to a familiar language we register various prosodic phenomena that speakers routinely produce. Second, we typically register stable vocal characteristics of speakers, such as pitch, tempo or accent, that are often systematically related to various properties of the speaker. Thus, the answer to the question of what we typically experience when listening to a familiar language is likely to be a complex one. Dedicated attention is needed to understand the nature and scope of phenomenology that pertains to linguistic communication. This paper lays some groundwork for that project.

Keywords: Linguistic understanding, experiences, prosody, linguistic communication.

* I would like to thank Nick Allott, David Pereplyotchik and Deirdre Wilson for detailed comments on this paper, as well as Barry Smith and the participants of the Philosophy of Language and Linguistics workshop in Dubrovnik in 2019 for helpful discussions on the topic. I am grateful to Dunja Jutronić for her help with preparing the draft. I would like to specially acknowledge the Institute of Philosophy in London, where I was a visiting researcher between 2018-2020. This work was supported by the Mobility Grant Fellowship Programme (FRICON) funded by The Research Council of Norway and the Marie Skłodowska-Curie Actions (project number: 275251).

1. *Introduction*

You are at the seaside leisurely flicking through a magazine. A friend calls out your name and asks if you would like to take another swim. There is energy and enthusiasm in her voice, and you have a strong impression that she is inviting you because she wants to take a swim as well. Your answer will probably be a quick cheerful 'yes'.

What do we experience when listening to a familiar language? For example, what do we experience when we hear a friend enthusiastically inviting us to take a swim? There is a *liberal* strategy for answering that question. After all, there are so many things one might experience when listening to an utterance produced in a familiar language. An endless variety of impressions may arise when we hear others speaking and communicate with them. Upon hearing your friend's invitation to take a swim, you might have a vivid memory of last year's summer holidays. An invitation might remind you of a scene from your favourite movie. You might feel relaxed and safe. You might feel threatened, as in the case when, unbeknownst to your friend, you nearly drowned during your last swim. You might have an unpleasant impression that your friend is nudging you to embrace a certain kind of healthy lifestyle, or you might have an impression that you and your friend have the same needs.

This would then be the answer provided by those who adopt the liberal strategy: virtually anything might be experienced when listening to a familiar language. However, the liberal strategy strikes me as an evasive one. Such varied experiences and impressions as those described above can, and perhaps often do arise when we listen to others speaking in a familiar language, but they *need* not. Moreover, when they do arise, they often do so with no specific regularity or order. A more fruitful way of addressing the question of what we experience when listening to a familiar language is to focus on elements that might be *systematically* experienced. When I ask what we experience when listening to a familiar language in this paper, I am asking about, and will consider, only those elements that typically and systematically arise in linguistic communication and could thus typically be part of our overall conscious experience.

How can we decide whether something is a candidate for being experienced *systematically*? The elements that I will consider in this paper are those that: (a) result from forms of expression and information transfer that are available to typically developed speakers, and (b) that typically developed hearers register and experience thanks to specific psychological and linguistic mechanisms employed in voice perception and spoken linguistic communication. Systematicity, as understood here, will not imply that such elements would need to be present in our experience on each and every occasion. For example, if a speaker does not convey or reveal some information when producing a linguistic utterance, a hearer will not register that information. Thus, I will not

go so far as to argue that the elements discussed in this paper would be necessary and sufficient for the case described. Linguistic communication and its phenomenology are complex phenomena and providing such conditions would be an overly ambitious task, at least in one go. Still, many real-life phenomena occur sufficiently frequently in normal conditions to warrant philosophical attention, even though exceptions may arise.

So what do we *systematically* experience in typical cases of hearing an utterance in a familiar language? One strategy would be to provide a quick and intuitive answer: we experience understanding an utterance or what the speaker said or communicated by uttering a sentence. When my friend asks me whether I would like to take a swim, I have an experience of her asking whether I would like to take a swim, or, in other words, I have an experience of understanding that she asked me whether I would like to take another swim and in that way invited me to do so.[1] This intuitive answer seems both quick and simple. In recent literature on the epistemology of linguistic understanding, experiences of what was said with an utterance, experiences of meanings communicated with utterances or experiences of understanding, depending on one's preferred terminology, have been a subject of intense debate. As we shall see (section 2) neither explaining the nature of such states nor explaining their epistemic roles is an easy task. This might be one reason why most discussions that might provide insights into the question of what we experience when listening to a familiar language have so far focused on experiences of meanings communicated with an utterance (or of linguistic understanding).

This focus may suggest, that the question about the phenomenology of linguistic communication boils down to whatever we can systematically say about the phenomenology of meanings or the phenomenology of understanding. In this paper, I will argue that this is not the case. The intuitive strategy described above would be too *restrictive* to provide a satisfying answer to the question of what we systematically experience when listening to linguistic utterances in a familiar language. Understanding or grasping a meaning conveyed by the speaker might of course be an important element of what we experience there and then. But in the course of this paper I will provide evidence that apart from this, two *other* elements that typically accompany the production of spoken linguistic utterances should be considered as candidates for what is systematically experienced when we listen to linguistic utterances in a familiar language.

First, I will argue that when we listen to a familiar language we also typically register a variety of *prosodic phenomena* that speakers routinely produce, in both controlled and spontaneous manner (Whar-

[1] In this and other cases, both the content of what was said with an utterance and its force, in this case an invitation, are commonly taken to be experienced (e.g. Fricker 2003).

ton 2009). Prosodic phenomena take many forms and often make important contributions to linguistic communication, affecting the everyday interactions involved. As such, they should be considered when addressing the question of what is systematically experienced by hearers. A friendly and enthusiastic invitation to take a swim sounds different from an indifferent one, or from one that merely attempts to sound friendly and enthusiastic.

Second, I will argue that, when we listen to linguistic utterances in a familiar language, we typically also register *stable vocal characteristics* of speakers, such as pitch and tempo, which are determined by the physiology of the speaker's vocal apparatus and the circumstances of vocal production, as well as vocal characteristics that result from sociolinguistic environment, such as accent. Importantly, such stable vocal characteristics are often systematically related to various properties of the speaker, such as sex, age, ethnicity (Belin et al. 2004, Baumann and Belin 2010; Owren et al. 2007; Mulac and Giles 1996; Rakić 2019). Thus, hearing a linguistic utterance in a particular voice will normally reveal a lot of important information about the speaker that the speaker does not intend to communicate.[2]

The evidence presented in this paper will show that the answer to the question of what we typically experience when listening to a familiar language is likely to be a complex one. The phenomenology of linguistic communication is probably richer than many mainstream philosophical debates on linguistic understanding would suggest. Careful investigation of whether and to what extent the three elements presented in the paper could enter heares' consciousness and what signalling functions they have is needed. This kind of investigation will require a detailed analysis of both philosophical and empirical arguments and goes beyond the scope of this paper. The goal of this paper is to lay some groundwork for the project of addressing the question of what we systematically experience when listening to a familiar language by pointing to new areas of research. For readers who are not immediately excited and curious about the question discussed in this paper, I would like to offer a brief explanation of why I think it is both interesting and important, thus making the task worthwhile.

First, it is interesting and important to describe and understand what we could systematically experience when listening to a familiar language. While a lot of philosophical attention has been paid to describing and explaining the nature of visual experiences and (somewhat less) to the nature of auditory experience of environmental sounds, apart from debates on experiences of understanding, there has been relatively less interest in the philosophical investigation of auditory experiences of listening to a familiar language (for rare ex-

[2] In this paper I focus entirely on spoken language and its phenomenology. Parallels and differences in the typical phenomenology of reading require separate discussion.

ceptions see: Smith 2009, Di Bona 2017, Drożdżowicz 2020). And yet such experiences are without a doubt part and parcel of our auditory environment, given that linguistic communication is a pervasive form of sharing information among humans. Second, systematic impressions of speakers can and often do play a role in linguistic communication by steering, affecting and sometimes biasing social interactions that rely on it. In the example above, specific properties of your friend's voice, e.g. those relating to the length of her vocal tract, determine how she sounds, namely as a middle-aged woman. Those properties allow you to recognize the voice that utters the invitation as the one belonging to your friend. In order to understand how we interact linguistically, dedicated attention is needed to provide an inclusive picture of what is experienced in linguistic communication.

The paper is structured into three main sections, each presenting one of the three elements that are typically registered when listening to an utterance in a familiar language. In section 2 I present recent debates concerning experiences of understanding what was said or communicated with an utterance. I summarize the main results concerning this topic and focus on pointing out questions for future research. I will briefly sketch my own view on the matter but will not defend it in detail here. In section 3 I present evidence which supports the claim that when we listen to a familiar language we commonly register a whole variety of prosodic phenomena and argue that they are another candidate to consider for what we could systematically experience in linguistic communication. Section 4 presents evidence for the claim that we also systematically register stable vocal characteristics of a speaker that reveal important information about them. Impressions based on such characteristics are yet another candidate for what we could systematically experience. I conclude in section 5.

2. *Experiencing linguistic understanding*

Setting aside cases of unsuccessful communication, when you hear an utterance in a familiar language, you typically come to understand *what the speaker communicated* with that utterance on a particular occasion. Many have argued that in such cases competent language users experience states of understanding linguistic utterances (Hunter 1998; Fricker 2003; Reiland 2015; Nes 2016; Brogaard 2018). Simple though it is, this observation has been a starting point for intense debates about the nature and epistemic roles of the experience in question.

Such experiences are commonly illustrated using so-called *contrast cases*. Imagine again hearing your friend utter the invitation "Would you like to take another swim?". Imagine now that everything is the same, except that your friend is speaking in a language totally unfamiliar to you, and says the same thing, i.e. invites you to take another swim. What will strike you is that the experience you have when listening to the second utterance in a language you do not know differs dra-

matically from the experience of a language you understand and speak fluently.[3] This observation is typically taken to be important *prima facie* (although not the only) evidence for the claim that experiences of linguistic understanding exist.[4]

Here is how such experiences are often portrayed in the current debate. It is generally agreed that when a hearer has an experience of understanding an utterance, she grasps at least one proposition, which would roughly correspond to the asserted meaning of that utterance. Most participants in this debate seem also to agree that typically such experiences would involve or somehow indicate grasping the enriched, saturated meaning of an utterance and not the minimal meaning (Fricker 2003; Nes 2016; Brogaard 2018; Gasparri and Murez 2019). For example, when you have an experience of understanding your friend uttering "Would you like to take another swim?", as in the above situation, you grasp, among other things, that: "you", in this context, refers to yourself, not to anyone else who might be at the beach, that the invitation, in this context, is to take a swim now or in the near future and not at just any point in the future, etc. Apart from rare occasions, hearers do not grasp minimal meanings with unassigned referents and unresolved ambiguities (Smith 2010).

When characterizing experiences of understanding linguistic utterances, it is usual to list their *involuntary* nature, *automaticity,* and prima facie *compelling* character. Normally we have little or no control over whether upon listening to an utterance in a familiar language we experience understanding it or not (Fodor 1983: 52). Such experiences are usually taken to arise spontaneously and automatically as soon as an utterance is heard. According to many, their immediate presence serves as *prima facie* compelling evidence for beliefs about what a speaker intended to communicate with an utterance (e.g. Fricker 2003; Brogaard 2018). Those features have been argued by some to support the idea that experiences of meanings or of linguistic understanding are interestingly similar to paradigmatic perceptual experiences (Hunter 1998; Bayne 2009; Siegel 2006; Brogaard 2018), but the exact nature of this similarity has been a subject of considerable debate.

At least three sets of questions have animated recent debates about experiences of linguistic understanding. These concern: (1) their nature, (2) their epistemic roles, and (3) the methods used to investigate them. Starting with (1), according to what has been described as the *semantic perceptual view,* properties like having meaning x can be represented in the hearer's perceptual experience (Siegel 2011; Bayne

[3] Moreover, typically utterances in unfamiliar languages are not perceived as strings of words, given that hearers are not sensitive to phonemes of a particular language (O'Callaghan 2011).

[4] This is a minimal commitment that many in this debate accept. Other evidence comes from the phenomenological shift that occurs when listening to sinewave speech (Remez et al., 1981). For a detailed discussion of this case and evidence it provides, see O'Callaghan (2011; 2015).

2009; Brogaard 2018, 2019). On that approach, the nature of experiences of utterance understanding is *perceptual* (or of perceptual seemings, Brogaard 2018). The experience we have when listening to a familiar language is an experience of *hearing meanings.* Arguably, the properties that are perceived in this case are many: Tim Bayne (2009) observes that we perceive "*both* (low-level) changes in phonological structure *and* (high-level) semantic properties".[5] In order to have an experience of understanding an utterance in a familiar language, a hearer has to perceive the phonetic and phonological properties of an utterance (O'Callaghan 2015). But it is a contentious matter whether in such cases meaning properties are also perceived, and in recent years the semantic perceptual view has been the subject of some criticism.

According to Casey O'Callaghan (2011), the contrast cases that are often presented in favour of the semantic perceptual view are best explained in terms of differences in experiencing low-level phonological properties of linguistic utterances and thus do not support claims about semantic perception. The view has also been criticized on epistemological grounds. Brendan Balcerak Jackson (2019) has recently argued that the claim that experiences of hearing meaning provide immediate justification for hearers' beliefs about what a speaker communicated with an utterance (e.g. Brogaard 2018) is unfounded. Another contentious issue has been whether the semantic perceptual view is psychologically realistic and could be made compatible with our best knowledge about the psychology of linguistic comprehension (Drożdżowicz 2019), as well as whether it can accommodate the systematic role of context and background knowledge in linguistic communication (Brogaard 2018, 2019; cf Gasparri and Murez 2019). A related contender in this debate, the view that experiences of linguistic understanding are instances of *cognitive* phenomenology characteristic of cognitive states (Strawson 2010; Siewert 1998; Dodd 2014), has also been subject to criticisms on parallel grounds (Prinz 2011; see also Montague 2017).[6]

A quite different approach to the nature of such experiences maintains that meanings (or thoughts) are not the salient contents of such experiences, but rather that what hearers experience is fluency of understanding. On my own view (Drożdżowicz forthcoming), experiences of understanding are *epistemic feelings of linguistic fluency* that result from evaluative monitoring processes.[7] There is extensive evidence that such processes are typically involved in utterance comprehension (Nozarri and Novick 2017; Pickering and Garrod 2013). The perceptual appearance of understanding experiences is, on this view, explained as resulting from the deployment of early-stage auditory processes of speech perception. Thus, on my proposal, experiences of linguistic understand-

[5] There might also be other properties (morphological, syntactic) to consider.

[6] This passage draws on section 2 from Drożdżowicz 2019.

[7] For the purpose of this paper I briefly mention my view as one of the contenders in this debate. I defend it in detail in another paper (Drożdżowicz forthcoming).

ing are first and foremost metacognitive feelings that reveal the degree of the success in comprehending an utterance. For example, when you hear a friend inviting you to take another swim, you might have an experience of understanding that amounts to a quick immediate signal indicating that you have successfully comprehended your friend's invitation and can immediately proceed to produce an answer, act on that invitation, etc. For these purposes, you do not have to represent or reconstruct the communicated meaning of her utterance as part of your conscious experience, but simply take for granted the feeling that you got the message of her utterance right, and act on it.[8] Although the outcomes of the ongoing debate on the nature of experiences of understanding still remain to be seen, the most discussed semantic perceptual view is currently under a lot of pressure. This opens up space for new contenders and calls for further investigation of question (1).

Questions concerning the epistemic roles of experiences of understanding (2), or of meanings conveyed with linguistic utterances, are of immediate interest to epistemologists working on linguistic communication and testimony. In recent debates on the epistemology of language understanding, it has been argued that such experiences: (a) justify beliefs about what a speaker communicated or said with an utterance (Hunter 1998; Brogaard 2018: 2969); (b) provide justification that is necessary for acquiring knowledge about what a speaker said (Fricker 2003: 345)[9]; (c) amount to what states of language understanding are (Pettit 2002: 544); and perhaps (d) trigger the "content entertaining" states of understanding (Longworth 2018: 825).[10]

Let us look at some evidence presented in favour of the claim that experiences of linguistic understanding provide important, justification for beliefs (Brogaard 2018) and/or knowledge about what the speaker conveyed with an utterance (Fricker 2003). Assuming that we are in a *typical* communicative context, i.e. one where both speaker and hearer are using the same language in a cooperative way (Fricker 2003: 332), what could be the epistemic contribution of an experience of understanding an utterance produced by the speaker? In our toy example, your friend asks you: "Would you like to take another swim?". Accord-

[8] This is compatible with the fact that in other, albeit less prevalent, cases of obstructed communication you might need to reflectively reconstruct the meaning you have grasped.

[9] Fricker (2003) uses the notion of what is said when describing such knowledge and beliefs, but her clarification of experiences of understanding (and corresponding beliefs) suggests that their contents concern not just what is strictly speaking said with an utterance, but what a speaker intended to communicate with an utterance on a particular occasion, leaving it open whether all pragmatic meanings (e.g. implicatures, metaphors) can be experienced in a similar manner. A similar notion is used in Brogaard (2018).

[10] Guy Longworth's notion of 'perceptual encounter with an utterance' seems parallel to the notion of experience of understanding. On the other hand, his content-entertaining states seem to involve both perceptual and belief-like elements, so they might perhaps be closer to the experiences discussed here.

ing to Fricker (2003) and Brogaard (2018), in this case your experience of understanding what she communicated with that utterance provides an immediate *prima facie* justification for your belief about what she said,[11] i.e. that she has invited you to take another swim. However, having a relevant experience of understanding would rarely suffice for your belief to be justified or to afford knowledge. On their view, you would also need to have secured a kind of warrant that is captured in broadly externalist or reliabilist terms. For example, one could maintain that for such a belief to be justified, a hearer must exercise reliably functioning linguistic capacities. Lucky beliefs, based on matching experiences of understanding produced by the unreliable workings of the language system, typically would not count as justified.

How then should we understand the strength of the claim that experiences of understanding are normally required for justified beliefs (and/or knowledge) about what a speaker said? Fricker's case of Ida (2003), initially presented as an argument against the reliabilist conception of language understanding (e.g. Schiffer 1987), is often discussed in this context:

> **IDA:** Ida has an internal, module-like device implemented in her brain that provides her with correct beliefs about the meanings of utterances in Russian: "When Ida hears a sentence in Russian, it sounds like meaningless noise to her. Yet after hearing it, she finds herself with a strong inclination to believe that a certain speech act has been effected by that – to her – meaningless burst of noise. Ida instantiates the correct 'template'. Her beliefs about what is said in Russian utterances are due to a language processing, belief-generating module in her, and are reliably true." (Fricker 2003: 337)

According to Fricker, without experiences of understanding Russian utterances, Ida cannot know what Russian speakers say. The example is used by Fricker to support her claim that "the phenomenology of understanding *is essential* to how knowledge of what is said is gained, in normal language use" (345). Although Ida satisfies the reliabilist conception of understanding, she does not have any reasons available to her to support her beliefs about what utterances in Russian mean. Fricker argues that, intuitively, Ida does not know what utterances in Russian mean. On the contrary, utterances in Russian sound meaningless to her. Ida's case is different from the case of normal language understanding because her way of gaining reliable beliefs about meaning is "phenomenally lacking".[12]

I would like to suggest that the epistemic benefit that typical language users seem to have over Ida in this case could perhaps be cap-

[11] In these discussions a distinction between beliefs about what a speaker said with an utterance and beliefs in the content of their utterance is typically assumed.

[12] As presented in this scenario, Ida is of course very different from typical language users who have a normally developed language system. The dialectic role of this example is therefore constrained by the stipulations about how Ida's module works.

tured by the notion of *doxastic* justification. A belief is doxastically justified to the extent that it is epistemically supported by the reasons on which the agent bases it (Dormandy 2018: 77). Although Ida's beliefs about what Russian speakers say are reliably causally sustained by the workings of her internal module, by their very definition, the workings of her module do not fulfil the condition for doxastic justification, i.e. they cannot be *treated as a reason* for her beliefs. But an observation-based meta-belief that she is reliably forming such beliefs due to the inner workings of her module could provide such a reason. One could argue that this would be an improvement on Ida's current epistemic situation. From a believer's perspective, obtaining any good reason can increase the durability and confidence of an agent's belief (Dormandy 2018). Cases like IDA should not lead us to conclude that experiences of understanding are strictly speaking necessary for justified belief and/or knowledge about what a speaker conveyed with an utterance. After all, there may be other possibly valuable ways of forming such beliefs. Consider the following hypothetical case:

> **ADA:** Ada has just met a Portuguese friend. She does not speak Portuguese. However, there is a language pill she could take that would make her acquire reliable true beliefs about the meanings of utterances in Portuguese. Ada would be instantiating a correct template from utterances in Portuguese to beliefs about what these utterances mean. There is only one potential downside - the pill does not induce the appropriate, typical phenomenology of understanding Portuguese utterances that most speakers of Portuguese typically enjoy. Should Ada take the pill?[13]

Intuitively, Ada would be better off if she took the pill, for she could then come to correctly and reliably believe what her Portuguese friend is saying. Even without the typical corresponding phenomenology of understanding, Ada would have *more understanding* of Portuguese than before taking the pill. Ada's beliefs about what Portuguese speakers say could be epistemically supported by her meta-belief that the pill allows her to reliably form such beliefs. Notwithstanding the intuitive verdict that Ada should take the pill, her epistemic situation would still be quite different from that of typical Portuguese speakers. There seem to be some epistemic benefits which would not be available to her, namely, an experience of understanding that could also doxastically support a corresponding belief.

Arguably, a somewhat different story about the epistemic role of experiences of understanding would accompany the view that they are epistemic feelings of linguistic fluency (Drożdżowicz forthcoming), since on that view such experiences do not present meanings as their contents but merely signal the fluency of their comprehension. Many epistemic feelings can be useful for deliberate metacognitive purposes, such as guiding a subject's attention or motivating one to reconsider

[13] I thank Sandy Goldberg for suggesting this example.

one's epistemic standing (Dokic 2012; Koriat 2007). Epistemic feelings of linguistic fluency could also fulfil some such roles. In a typical case, a feeling of fluency could signal that the hearer can proceed to utilize information about an utterance in communication, belief formation, and action. On the other hand, an epistemic feeling indicating a lower level of fluency may signal a need to allocate more resources, repair, etc. Epistemic feelings of linguistic fluency could guide our cognitive functioning in several important ways and in this way fulfil some epistemic roles.[14] Unsurprisingly, at least some answers to questions about the epistemic roles of such states (2) seem to depend on our views of their nature (1).

The final set of questions (3), concerns methodological issues about how the nature of experiences of understanding should be investigated. Which considerations should bear on the above questions and shape our views? Some people investigate the phenomenology of such experiences and build arguments based on contrast cases (e.g. Siegel 2010; Dodd 2014); others emphasize their epistemic roles in acquiring information and social interactions (e.g. Brogaard 2018; Balcerak Jackson 2019); still others advocate drawing on empirical evidence from psychological research on speech and utterance comprehension (e.g. O'Callaghan 2011; Gasparri and Murez 2020). Another complexity concerns whether and to what extent the philosophers' notion of conscious experience, as used in the debates on linguistic understanding, can be made compatible with currently available research on consciousness in psychology and neuroscience and, indeed, whether we currently have robust empirical evidence that could be informative for some of these issues. The intuitive answer to the question of what we experience when listening to a familiar language is only apparently a simple one, as the complex landscape of questions and views charted in this section illustrates. Issues concerning the phenomenology of linguistic understanding require more attention, given our common reliance on information acquired through linguistic communication (e.g. Goldberg 2018). But, as I will show in the next two sections, they do not exhaust what we should consider when we investigate the question of what we systematically experience when listening to a familiar language.

3. *Registering vocal prosody*

In many typical cases of linguistic communication, speakers produce linguistic utterances in a broader behavioural context. Spoken language is accompanied by a wide variety of non-verbal phenomena including vocal, facial, and bodily gestures (Wharton 2009). Usually, such gestures indicate the speaker's internal mental state—i.e., they convey information about their emotions, feelings, and attitudes toward the meanings expressed. Although we are often aware of using

[14] This passage draws on material from (Drożdżowicz forthcoming).

such gestures while speaking, and sometimes may even intentionally exploit them to achieve certain effects in our audience, in many cases, such gestures are produced spontaneously and beyond our conscious control. Non-linguistic gestures of various kinds can influence linguistic communication and may impact our understanding of utterances (Wharton 2009).

Linguistic communication is multimodal: it commonly exploits not only words, but also non-verbal *vocal* cues, as well as a whole variety of *visual* cues from the speaker's facial expression and bodily gestures. Bearing in mind that vocal and visual cues often interact, a phenomenon that recently has been a subject of intense study (e.g. Zhang et al. 2018; Frohlich et al. 2019), I will focus here on *prosody*—i.e., *vocal* elements of speech that are not individual phonetic segments (vowels and consonants) but properties of syllables and larger units of speech that commonly accompany the production of linguistic utterances (Speer and Blodgett 2006). This abstraction from visual input to linguistic communication is both necessary and warranted. The question to be investigated is what we systematically experience when listening to a familiar language. In addressing this, my focus will therefore be primarily on the auditory modality. In this section, I will provide evidence in support of the claim that prosodic phenomena are an important element that we routinely register when listening to a familiar language. Because of that they constitute a plausible candidate to consider for what can be systematically experienced in linguistic communication. Prosody can be and often is recognized without any visual input thanks to a specific psychological system that has been studied separately in experimental psychology of language. To illustrate, when your friend enthusiastically invites you to take another swim, in order to hear enthusiasm in her voice, you do not need to lift your sunhat to see her face or other bodily gestures. Whether she is smiling and vigorously imitating a crawl stroke, her enthusiasm, when expressed in the vocal prosody that accompanies her utterance, can be independently recognized.[15]

Prosody is an umbrella term encompassing a variety of vocal phenomena occurring in speech production. Specifically, it covers suprasegmental phonetic phenomena, i.e., properties that belong to larger units than phonemes, including syllables, phones, words, various intonation phrases and utterances (Speer and Blodgett 2006). There is general agreement that prosodic contributions to linguistic communication range from the intentionally produced, properly linguistic, and often language-specific ones (e.g., lexical tone, stress or pitch accent) to spontaneous, involuntary, or 'natural' ones (e.g. an angry, agitated or enthusiastic tone of voice) (Gussenhoven 2002; Pell 2002; Wharton

[15] The facial and bodily gestures, when perceived, may of course reinforce or modify your experience. An invitation produced in an angry tone of voice, but with a smile on the face would have a different effect than the one produced in a happy tone of voice and with a smile. Due to limited space, I leave discussion of such cases for another occasion.

2009: 139).[16] It is also commonly accepted that many contributions that prosodic gestures make to linguistic communication are context-dependent (Hirschberg 2002). Context may determine the degree of their contribution (hearing enthusiasm in your friend's voice may for example depend on your expectations about how much in general she likes to swim). It may also entirely determine the nature of a specific contribution (hearing enthusiasm in your friend's voice may strike you as fake and incongruent with the invitation, given that on such occasions she almost always speaks in a fairly neutral, flat tone of voice). Because of that, the prospects for a simple mapping from many prosodic phenomena to their communicative contributions are generally agreed to be dim (Hirschberg 2002, see also Wharton 2009: ch. 6).

The category of intentionally produced *properly linguistic prosodic gestures* is wide and includes, among other things, phenomena such as: contour variation, variation in location and type of pitch accents (e.g. nuclear stress on a single word), accent on discourse markers (e.g. but, although), accent on new information as opposed to what is given, accent on focus-sensitive operators (e.g. *only*, *some*, *must*), phrasing variation to chunk information in an utterance, variation in timing and pitch range to mark speaker involvement, final lowering (see Hirschberg 2002). The spontaneous, 'natural' prosodic contributions overlap to a large degree with what has been investigated under the label of *emotional prosody*.

'Emotional prosody' is a term used to describe phenomena in which speakers communicate emotions, either unintentionally or intentionally, by modifying acoustic attributes of their voice, and how these vocal cues are perceived and recognized by listeners (Pell and Kotz 2011). It has been argued that the neurocognitive system responsible for the processing of emotional prosody in hearers is distinct from the system responsible for the processing of speech sounds (Pell 2006), as well as from systems responsible for the processing of socially-relevant information recovered from the voice, such as age or gender (Belin et al. 2004; Spreckelmeyer et al. 2009). Some evidence suggests that vocal expressions of emotions are perceived categorically (e.g. Laukka 2005), thereby corresponding to a set of basic human emotions that also have discrete forms of expression in the face (Ekman 1994, but see Barett 2017; Celeghin et al. 2017). Furthermore, vocal expressions of anger, disgust, fear, sadness, and happiness/joy can be accurately recognized when listening to a foreign language (e.g. Pell et al. 2009; Sherer et al. 2001). This suggests that at least these emotions have discrete acoustic-perceptual properties in the voice which manifest in similar ways across languages. According to empirical studies, vocally expressed emotions in speech are registered implicitly and automatically based on specific vocal cues (Kotz and Paulmann 2007). Studies also suggest

[16] Several distinct and possibly overlapping distinctions are grouped here following recent discussions on the topic.

that there are important differences in the underlying time course for typical recognition of basic emotions from vocal expressions. Anger, sadness, fear, and neutral expressions are recognized more accurately shortly after hearing vocal cues than happiness and disgust. However, as speech unfolds with time, recognition of happiness improves significantly towards the end of the utterance, while fear is recognized more accurately than other emotions (Pell and Kotz 2011).

Described in this way, emotional prosody is a particularly important channel of information about the speaker's mental state and is often a subtle but permanent aspect of what we register when listening to linguistic utterances in a familiar language (and as the above studies suggest, when we listen to foreign languages too). When your friend invites you to take another swim, a particular shade of happiness that I have labelled as 'enthusiasm' reveals how she feels about the prospects of going for a swim with you and colours her invitation in a subtle but important way. You are sensitive to that colouring, and recognize the emotional expression in her voice.

Prosodic cues of various types tend to create impressions and convey information about speakers' emotions or attitudes, rather than expressing full propositions or concepts in their own right, as words and utterances usually do (Wharton 2009: 141). But there is no doubt that prosody can and often does aid linguistic communication (Hirschberg 2002; Fodor 2002; Wharton 2009). Prosodic contributions to linguistic communication are something that hearers regularly and systematically draw on in linguistic interactions. For example, the specific emotional prosody that accompanies your friend's invitation may influence what in the end you will take her to be communicating:

(swim 1) Would you like to take another swim? (*happy, enthusiastic tone of voice*)
(swim 2) Would you like to take another swim? (*neutral tone of voice*)

When uttered in a happy, enthusiastic tone of voice (swim 1), the invitation, given certain contextual expectations that you have about your friend and her interest in swimming, will likely be reinforced by the accompanying emotional prosody. When uttered in a neutral tone of voice (swim 2), the invitation to swim may, given some contextual assumptions, indicate that your friend is not, after all, excited about the prospects of your taking another swim. Perhaps she is offering it out of politeness, knowing that you love to swim but are afraid of doing it alone, etc. Registering the emotional prosody in your friend's voice will guide your overall interpretation of what happens in this linguistic interaction.

Another common example of how prosody impacts linguistic communication is that of intentionally employing a specific type of *stress pattern*:

(swim 3) Would you like to take another **swim**?
(swim 4) Would **you** like to take **another** swim?

When uttered with a neutral stress pattern, where the nuclear pitch accent falls at the end of the utterance, and specifically, on the last word (swim 3), given certain contextual expectations, you have grounds for taking your friend's invitation at face value. When uttered with a contrastive stress pattern, with an accent on 'you' and 'another' (swim 4), your friend's utterance may strike you as indicating something different from an invitation to take another swim. With this contrastive stress pattern, your friend may be indicating that she would be surprised if you accepted the invitation or that she doubts your stamina. Whichever interpretation might be most likely in this context, there is a clear sense in which prosodic variation in the stress pattern that accompanies an utterance contributes to what you will ultimately get from this linguistic interaction. It will affect how you interpret the invitation and probably also how you respond to it.[17]

In typical linguistic interactions, when we listen to a familiar language, we might have an experience of understanding what the speaker communicated with an utterance. But as the above evidence and examples show, we also routinely register the prosody that accompanies and partly constitutes linguistic utterances. We draw on information conveyed by vocal cues that are produced by the speaker both in a spontaneous and an intentional manner. Speakers have the resources to produce prosodic phenomena and do so regularly. Hearers register and draw on prosodic phenomena thanks to specific psychological mechanisms.

The above observations have led many linguists to ask whether prosody, given its contribution to linguistic interactions, may encode some relatively stable meanings, and if so what kind of meanings those could be. As already mentioned, any claims about prosodic meanings or prosodic code are bound to be limited by the overwhelming context-dependence of the contributions that prosodic information makes (Hirschberg 2002; Wharton 2009). According to Gussenhoven and colleagues (Gussenhoven 2002; Chen and Gussenhoven 2003), our understanding of various prosodic gestures is governed by both biological and properly linguistic codes. What they call the effort code is a biological code that connects the amount of energy that speakers utilize in speech production with specific prosodic cues to a range of interpretive effects. "An increase in effort may lead to increased articulatory precision, creating an impression of 'helpfulness', or 'obligingness'; or it may result in a wider pitch range, creating an impression of 'forcefulness' or 'certainty' or conveying affective meanings such as 'agitation' or 'surprise'" (in Wharton 2009: 143). A different approach to explaining the communicative contributions of prosodic gestures can be found in Wharton (2009). In his view, both spontaneous, uncontrolled and intentionally produced, properly linguistic prosodic gestures might encode *procedural* information, i.e. information where a word (or other linguistic

[17] For other interesting examples see Wilson and Wharton 2006.

device) encodes information specifically geared to guiding the hearer during the inferential phase of comprehension (145). In this sense, he argues, many prosodic gestures can be seen as encoding procedural meaning (Blakemore 2002; Escandell et al. 2011).

For the purposes of this paper, I will not take a stance in the debate about whether and how we could model the meaningful contributions that prosody makes to linguistic communication. It suffices to say that prosody is an important element of spoken utterances that influences many linguistic interactions. Prosody is thus a plausible candidate to consider when investigating what we systematically experience when listening to a familiar language. Whether and which aspects of prosodic information could actually surface to hearers' consciousness should be carefully investigated with both philosophical and empirical tools. In recent years prosody has received a lot of attention in theoretical linguistics and in experimental psycholinguistics. Much work remains to be done, and new research avenues emerged, such as the use of prosody in artificial text-to-speech and speech-understanding systems (Hirschberg 2002). Yet, curiously, prosody is rarely noticed in mainstream philosophical discussions of linguistic communication and understanding. This is surprising, given the abundance of evidence for the claim that prosody is commonly produced by speakers, and routinely registered by hearers. Thus, a full answer to the question of what we experience when listening to a familiar language requires an empirically-informed account of the role of prosodic cues.

One might ask whether, in light of the evidence above, a distinction between possible contributions of linguistic understanding and prosody to what we experience would be in fact warranted. Prosodic cues can affect utterance understanding. The contrastive stress pattern can influence even the truth-conditions of an utterance (as in "Sue only *spoke* to Laura, vs Sue only spoke *to Laura*"), thereby affecting what meaning or proposition we grasp upon hearing it.[18] Neither allowing for prosodic contributions to utterance interpretation, nor for interactions between communicated meanings and prosody, would I think undermine the idea that we might be able to experience prosody as something different from understanding an utterance. In many cases it makes sense to distinguish between understanding an utterance and hearing the accompanying prosody.

First, in many cases there is an intuitive sense in which we seem to register understanding an utterance and its vocal prosodic material separately. This is why we can capture our understanding by paraphrasing the sentence uttered, as well as capture the prosodic characteristics of the utterance by, for example, noting the emotions conveyed by a speaker's voice or the stress pattern used. Second, we have psycholinguistic evidence that the processing of verbal speech material and prosodic speech material is performed by two largely independent

[18] I thank Deirdre Wilson for helpful comments regarding this issue.

systems (e.g. Pell 2006; Belin et al. 2004). Third, verbal and prosodic contributions can come apart: a neutral tone of voice may not aid interpretation at all, and emotional prosody may be detected even when the phonological sounds and word meanings are not, as in the case of hearing emotions in an utterance produced in a foreign language or, as some studies suggest, in an artificial meaningless speech signal (Grandjean et al. 2005). Although linguistic understanding and prosody may and often do come together in linguistic communication, their possible contributions to what we experience when listening to a familiar language can be considered (at least to some degree) separately.

4. *Registering stable vocal characteristics of a speaker*

In this section, I argue that in typical cases of listening to utterances produced in a familiar language, we also typically detect stable vocal characteristics of a speaker and that those are another candidate to consider for what could systematically experience when listening to a familiar language. Usually, such vocal characteristics are not used for the purpose of and do not aid linguistic communication, but nevertheless reveal a lot of interesting and important information about the speaker. There are certain vocal characteristics of a speaker's voice that she or he cannot easily conceal when producing linguistic utterances. Hearers are sensitive to those characteristics and register them when listening to speech from a particular speaker. Many vocal parameters that are exhibited in vocal production coming from a particular voice systematically correspond to and indicate important properties of a speaker. Voice conveys not only rich information about a speaker's emotional state and attitudes (Pell and Kotz 2011; Bänzigier et al. 2014), as explained in section 3, but also provides extra-linguistic cues that reflect more stable speaker properties, including: identity (Baumann and Belin 2010), biological sex (Owren et al. 2007), age (Mulac and Giles 1996), and even the socioeconomic status and regional background of a speaker (e.g. Rakić 2019).

How is it possible that our voices can reveal so much about us? Vocal sounds are generated by the interaction of a source (the vocal folds in the larynx) and a filter, i.e. the vocal tract above the larynx (Ghazanfar and Rendall 2008). Voiced sounds correspond to a periodic oscillation of the vocal folds with a well-defined *fundamental frequency* (f0). Although for an individual speaker the range of f0 values can vary quite a lot during normal phonation or singing, the average f0 of a particular speaker is to a large extent a function of the size of the vocal folds (Latinus and Belin 2011). Male vocal folds tend to be longer and thicker than female vocal folds, causing them to vibrate at approximately half the frequency (100–120 Hz) than average female vocal folds (200–240 Hz) (Stevens 1998). This is why female and male voices tend to differ systematically in a way that is often recognized by hearers. Human voices tend to vary extensively. Small differences in the dimensions and histology of the

individual body parts that speakers use in phonation result in great individual variability among speakers in the individual acoustic patterns they can produce. Interindividual differences in the dimensions of the vocal folds and their tension during speech production cause variation in mean fundamental frequency (f0) and voice quality. Differences in other parts of the vocal tract, such as the nasal passage, result in differences in the absolute and relative positions of the resonance frequencies of the vocal tract (for details see Schweinberger et al. 2014).

Given such a multitude of factors determining how our voices sound, it is important to understand which of the perceptible vocal characteristics are utilized by hearers to identify speakers' voices and differentiate between them. Several studies suggest that perception of the fundamental frequency of a speaker's voice is a key parameter in recognizing the voice (e.g., Bauman and Belin 2010). Other studies reveal a more complex picture, where other parameters such as jitter (local variations in period length), shimmer (local variation in period amplitude) and harmonics-to-noise ratio are also utilized in voice perception (e.g., Kreiman and Sidtis 2011). It is generally agreed that the human perceptual ability to recognize voices is realized by a particular neurocognitive system. Neuroimaging studies have identified several brain areas in the temporal cortex, located in the middle parts of the superior temporal sulcus (STS) bilaterally, which show a particular sensitivity to voices, irrespective of whether they contain speech (Belin et al. 2004; Binder et al. 2000). Moreover, there is evidence that information about the stable vocal characteristics of a speaker is processed largely independently of the prosodic vocal information described in section 3. Studies investigating the perception of affective prosody show a greater activation of the right temporal lobe and right inferior prefrontal cortex (Mitchell et al. 2003) for prosodic information.[19] Other studies (e.g. Belin and Zatorre 2003) seem to confirm the role of anterior temporal lobe regions of the right hemisphere, particularly right anterior STS regions, in processing information about speakers' voices related to their identity.[20]

For humans, voices are among the most prevalent and salient sounds in the auditory environment. Our ability to analyze the information that is contained in voices is important for many social interactions. Take our swimming example again. Even without seeing your friend approaching you at the seaside, when you hear her enthusiastically uttering the sentence with which she invites you to take another swim, you would normally immediately recognize the voice you hear as the one that belongs to your friend. When you hear her inviting you

[19] The perception of identity information in the voice has been examined in several neuroimaging studies suggesting that the anterior temporal lobes in both hemispheres are more active during speaker identification than during emotion identification (Imaizumi et al., 1997).

[20] Passages on pages 381 and 382 up till this point draw on material from section 2 of Drożdżowicz 2020.

to take another swim, you register information about the specific vocal parameters of her voice, such as its fundamental frequency, tempo, and the resonance frequencies that are determined by the anatomy of your friend's vocal folds and vocal tract. It is by registering these parameters that you can hear the voice that utters the invitation as belonging to a middle-aged female. Moreover, given your familiarity with your friend's voice, upon hearing those vocal parameters, you immediately recognize the voice as *hers*. Hearing the same invitation to take a swim uttered in the same context but in a voice that does not belong to your friend will result in a markedly different auditory experience of the stable vocal characteristics. Needless to say, an invitation from a stranger will have a very different communicative effect. Our sensitivity to human voices is an amazing perceptual advantage that allows us to effortlessly and typically accurately track a source of spoken linguistic utterances and in this way facilitates interactions based on linguistic communication.

As already mentioned, there are other properties of speakers that are systematically indicated by relatively stable vocal characteristics of a speaker's voice. Among them, regional dialect and foreign accent are properties of speakers that we are sensitive to. At this point you may no longer pay attention to her accent, but the first time you heard your friend speaking, it may have struck you that she speaks with a slight Danish accent. Having watched multiple Danish crime series, you were actually able to correctly identify her accent as Danish, though many of your friends initially struggled with that. In many cases, we recognize stable vocal characteristics of a speaker that point to an identifiable regional dialect or foreign accent. As we detect them, we can often become aware of the speaker's place of origin, ethnic background, and sometimes even their socioeconomic status. It is not only what we say, but also how we sound, that has a power to generate impressions beyond what we intend to convey and often beyond what we would like to reveal to interlocutors.

Hearers' impressions of speaker properties based on their perception of stable vocal characteristics need not always facilitate linguistic communication or the social interactions that draw on it. Take foreign accent as an example. Foreign accent can influence social interactions based on linguistic communication. Leaving aside rare exceptions in which foreign accents are perceived positively (Gibson et al 2017), there is evidence suggesting that in various sociolinguistic contexts speakers with foreign accents are judged to be less intelligent, less trustworthy, less educated and less competent than native speakers (e.g. Dewaele and McCloskey 2015; Dragojevic et al. 2016; Fraser and Kelly 2012; Fuertes et al. 2012; Gluszek and Dovidio 2010; Livingston et al. 2017). Negative bias towards foreign-accepted speech is present from early childhood (Kinzler et al. 2007). At the age of 11, children tend to trust native-accented speakers more (Kinzler et al. 2011).

The foreign accent bias may have (at least) two origins. One is linguistic: foreign accent may decrease 'processing fluency' and lead to

lower intelligibility of the speaker (Deterding and Kirkpatrick 2006; Cristia et al. 2012).[21] The other is social: foreign-accented speakers are rapidly categorised as out-group members. In this way, foreign accent may lead to negative evaluation of speaker's competence by being a function of shared negative attitudes towards the ethnicity of the accented speaker (Lippi-Green 1997; Roessel et al. 2019). The foreign accent bias has been shown to lead to discrimination in various contexts, for example in the courtroom (Solan and Tiersma 2004) and in job interviews (Huang 2013; Hansen and Dovidio 2016).

Although, your friend's Danish accent may not be an issue when she is inviting you to take a swim, information about the speaker's stable vocal characteristics can systematically influence social interactions. Registering stable vocal characteristics of a speaker and the resulting impressions may facilitate linguistic communication and social interactions, but it may also systematically impede them. Such vocal characteristics are another candidate to consider when investigating what we systematically experience in linguistic communication. Our sensitivity to stable vocal characteristics of a speaker raises interesting ethical questions about linguistic interactions that are affected by our impressions of the speaker. It also invites us to consider whether and which properties of vocal production could be perceived by hearers and which might result from inference and the underlying implicit beliefs that hearers have about speakers. This is a complex question that requires detailed treatment. Where and how exactly the border between auditory vocal perception and audition-based cognition of speaker properties is to be drawn is a difficult matter that is likely to generate an intense discussion and requires both philosophical and empirical investigation (e.g. Di Bona 2017).

5. *Concluding remarks*

I have argued that at least three elements need to be considered when we ask what we systematically experience when listening to a familiar language. (i) We perceive speech sounds and typically seem to have an experience of *understanding* what the speaker communicated with an utterance on a particular occasion. (ii) We register various forms of *prosody* and thanks to that we can learn about speaker's attitudes and mental states indicated by prosodic cues. (iii) We register the speaker's *stable vocal characteristics* and have systematic impressions about the speaker's identity. Those *three elements* should be investigated when we consider the question of what could be systematically present in our experience when we listen to a familiar language, given that they:

[21] This might involve a pragmatic component: in some contexts foreign accents might increase processing effort for native speakers when first encountered, and might therefore affect assessments of relevance, and ultimately competence etc. This effect could in some cases diminish as hearers become more familiar with the accent and it becomes easier to process. I thank Deirdre Wilson for these points.

(a) result from common forms of linguistic expression and information transfer in humans, and (b) are registered by hearers thanks to the specific psychological mechanisms that are employed in spoken linguistic communication and voice perception.

As already mentioned, the three elements need not *always* be present when we listen to linguistic utterances. An utterance may not be understood, a speaker may produce an utterance in a flat tone, or they may have a voice that makes it particularly difficult to identify any speaker properties. Typically, however, the three elements are routinely produced and registered in spoken linguistic communication. How strong is the claim about the *systematic* presence of these three elements? This type of systematicity is contingent on our biological and cultural evolution. It is not entirely impossible that there could exist forms of human linguistic communication that do not make use of prosody,[22] and we can even imagine speakers who lack typical vocal characteristics. Human speakers however, at least for now, communicate using their vocal apparatus that has evolved in a particular manner (Belin et al. 2004). Moreover, they tend to take advantage of prosodic forms of expression when producing linguistic utterances. Human hearers are sensitive to such stable vocal characteristics and prosodic phenomena.

The paper provides some preliminary work for addressing the question of what we systematically experience when listening to a familiar language. I believe that a more inclusive approach to address the question is required if we are to make progress on epistemic and moral questions concerning testimony and other forms of social interactions that draw on linguistic communication in different contexts.

References

Bayne, T. 2009. "Perception and the reach of phenomenal content." *Philosophical Quarterly* 59 (236).

Bänziger, T. et al. 2014. "The role of perceived voice and in emotion communication." *Journal of nonverbal behavior* 38 (1): 31–52.

Barrett, L. F. 2017. "The theory of constructed emotion: an active inference account of interoception and categorization." *Social cognitive and affective neuroscience* 12 (1): 1–23.

Baumann, O. and Belin, P. 2010. "Perceptual scaling of voice identity." *Psychological Research PRPF* 74 (1): 110.

Belin, P., and Zatorre, R. J. 2003. "Adaptation to the speaker's voice in the right anterior temporal lobe." *Neuroreport* 14 (16): 2105–2109.

Belin, P., Fecteau, S., and Bedard, C. 2004. "Thinking the voice: neural correlates of voice perception." *Trends in cognitive sciences* 8 (3): 129–135.

[22] This is of course the case for many artificial speakers, such as recorded and computer generated announcements where the prosody is flat or wrong and does not serve any communicative function (Hirschberg 2002). Moreover, in the literature on autism it is often claimed that emotional prosody is commonly impaired, both in production and comprehension, of speakers with ASD (Chevallier et al. 2011).

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., and Possing, E. T. 2000. "Human temporal lobe activation by speech and nonspeech sounds." *Cerebral cortex* 10 (5): 512–528.

Blakemore, D. 2002. *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers* (Vol. 99). Cambridge: Cambridge University Press.

Brogaard, B. 2018. "In defense of hearing meanings." *Synthese* 195 (7): 2967–2983.

Chen, A. and Gussenhoven, C. (2003). "Language-dependence in the signalling of attitude in speech." In *CHI 2003 conference on Human and Computer Interaction*.

Chevallier, C., Noveck, I., Happé, F. and Wilson, D. 2011. "What's in a voice? Prosody as a test case for the Theory of Mind account of autism." *Neuropsychologia* 49 (3): 507–517.

Celeghin, A., Diano, M., Bagnis, A., Viola, M. and Tamietto, M. 2017. "Basic emotions in human neuroscience: neuroimaging and beyond." *Frontiers in Psychology* 8: 1432.

Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A. and Floccia, C. 2012. "Linguistic processing of accented speech across the lifespan." *Frontiers in psychology* 3: 479.

Deterding, D. and Kirkpatrick, A. 2006. "Emerging south–east Asian Englishes and intelligibility." *World Englishes* 25 (3–4): 391–409.

Di Bona, E. 2017. "Towards a rich view of auditory experience." *Philosophical Studies* 174 (11): 2629–2643.

Dodd, J. 2014. "Realism and anti-realism about experiences of understanding." *Philosophical Studies* 168 (3).

Dokic, J. 2012. "Seeds of self-knowledge: noetic feelings and metacognition." In *Foundations of metacognition*, Chapter 6: 302–321.

Dormandy, K. 2018. "Epistemic Authority: Preemption or Proper Basing?" *Erkenntnis* 83 (4): 773–791.

Dragojevic, M., Giles, H., Beck, A. C. and Tatum, N. T. 2017. "The fluency principle: Why foreign accent strength negatively biases language attitudes." *Communication monographs* 84 (3): 385–405.

Drożdżowicz, A. 2019. "Do we hear meanings? – between perception and cognition." *Inquiry* 1–33.

Drożdżowicz, A. 2020. "Bringing back the voice – on the auditory objects of speech perception." *Synthese* 1–27.

Drożdżowicz, A. Forthcoming. "Experiences of understanding as epistemic feelings." *Mind & Language*.

Ekman, P. 1994. "All emotions are basic." *The nature of emotion: Fundamental questions* 15–19.

Escandell-Vidal, V., Leonetti, M. and Ahern, A. (eds.). 2011. *Procedural meaning: Problems and perspectives* (Vol. 25). Leiden: Brill.

Fodor, J. A. 1983. *The modularity of mind: An essay on faculty psychology*. Cambridge: MIT Press.

Fodor, J. D. 2002. "Psycholinguistics cannot escape prosody." In *Speech Prosody 2002, International Conference*.

Fraser, C. and Kelly, B. F. 2012. "Listening between the lines: Social assumptions around foreign accents." *Australian Review of Applied Linguistics* 35 (1): 74–93.

Fricker, E. 2003. "Understanding and knowledge of what is said." In A. Barber. *Epistemology of Language*. Oxford: Oxford University Press: 325–66

Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T. and van Schaik, C. P. 2019. "Multimodal communication and language origins: integrating gestures and vocalizations." *Biological Reviews* 94 (5): 1809–1829.

Fuertes, J. N., Gottdiener, W. H., Martin, H., Gilbert, T. C. and Giles, H. 2012. "A meta-analysis of the effects of speakers' accents on interpersonal evaluations." *European Journal of Social Psychology* 42 (1): 120–133.

Gasparri, L. and Murez, M. 2019. "Hearing meanings: the revenge of context." *Synthese* 1–24.

Ghazanfar, A. A. and Rendall, D. 2008. "Evolution of human vocal production." *Current Biology* 18 (11): R457–R460.

Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B. and Fedorenko, E. (2017). "Don't underestimate the benefits of being misunderstood." *Psychological science* 28 (6): 703–712.

Gluszek, A. and Dovidio, J. F. 2010. "The way they speak: A social psychological perspective on the stigma of nonnative accents in communication." *Personality and social psychology review* 14 (2): 214–237.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R. and Vuilleumier, P. 2005. "The voices of wrath: brain responses to angry prosody in meaningless speech." *Nature neuroscience* 8 (2): 145–146.

Gussenhoven, C. 2002. "Phonology of intonation." *Glot International* 6 (9/10): 271–284.

Hansen, K. and Dovidio, J. F. 2016. "Social dominance orientation, nonnative accents, and hiring recommendations." *Cultural Diversity and Ethnic Minority Psychology* 22 (4): 544.

Hirschberg, J. 2002. "Communication and prosody: Functional aspects of prosody." *Speech Communication* 36 (1–2): 31–43.

Huang, B. H. 2013. "The effects of accent familiarity and language teaching experience on raters' judgments of non-native speech." *System* 41 (3): 770–785.

Hunter, D. 1998. "Understanding and belief." *Philosophy and Phenomenological Research* 58 (3).

Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., Fukuda, H. ... and Kojima, S. 1997. "Vocal identification of speaker and emotion activates different brain regions." *Neuroreport* 8 (12): 2809–2812.

Jackson, B. B. 2019. "Against the perceptual model of utterance comprehension." *Philosophical Studies* 176 (2): 387–405.

Kinzler, K. D., Corriveau, K. H. and Harris, P. L. 2011. "Children's selective trust in native–accented speakers." *Developmental science* 14 (1): 106–111.

Koriat, A. 2007. "Metacognition and consciousness." In P. D. Zelazo, M. Moscovitch and E. Thompson (eds.). *The Cambridge Handbook of Consciousness*. Cambridge: Cambridge University Press: 289–326.

Kotz, S. A. and Paulmann, S. 2007. "When emotional prosody and semantics dance cheek to cheek: ERP evidence." *Brain research* 1151: 107–118.

Kreiman, J. and Sidtis, D. 2011. *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. Hoboken: John Wiley and Sons.

Latinus, M. and Belin, P. 2011. "Human voice perception." *Current Biology* 21 (4): R143–R145.

Laukka, P. 2005. "Categorical perception of vocal emotion expressions." *Emotion* 5 (3): 277.

Lev-Ari, S. 2015. "Comprehending non-native speakers: theory and evidence for adjustment in manner of processing." *Frontiers in psychology* 5: 1546.

Lev-Ari, S. and Keysar, B. 2010. "Why don't we believe non-native speakers? The influence of accent on credibility." *Journal of experimental social psychology* 46 (6): 1093–1096.

Lippi, R., Donati, S., Lippi-Green, R. and Donati, R. 1997. *English with an accent: Language, ideology, and discrimination in the United States*. Psychology Press.

Livingston, B. A., Schilpzand, P. and Erez, A. 2017. "Not what you expected to hear: Accented messages and their effect on choice." *Journal of Management* 43 (3): 804–833.

Longworth, G. 2018. "Understanding what was said." *Synthese* 195 (2): 815–834.

Mitchell, R. L., Elliott, R., Barry, M., Cruttenden, A. and Woodruff, P. W. 2003. "The neural response to emotional prosody, as revealed by functional magnetic resonance imaging." *Neuropsychologia* 41 (10): 1410–1421.

Montague, M. 2017. "Perception and cognitive phenomenology." *Philosophical Studies* 174 (8): 2045–2062.

Mulac, A. and Giles, H. 1996. "'You're Only As Old As You Sound'". *Health Communication* 8 (3): 199–215.

Nes, A. 2016. "On what we experience when we speak." *Phenomenology and Mind* 10: 58–85.

Nozari, N. and Novick, J. 2017. "Monitoring and control in language production." *Current Directions in Psychological Science* 26 (5): 403–410.

O'Callaghan, C. 2011. "Against hearing meanings." *The Philosophical Quarterly* 61 (245): 783–807.

O'Callaghan, C. 2015. "Speech Perception." In *Handbook of the Philosophy of Perception*. Oxford: Oxford University Press.

Owren, M. et al. 2007. "Listeners judge talker sex more efficiently." *Perception and psychophysics* 69 (6): 930–941.

Pell, M. D. 2002. "Surveying emotional prosody in the brain." In *Speech Prosody 2002, International Conference*: 77–82.

Pell, M. D. 2006. "Cerebral mechanisms for understanding emotional prosody in speech." *Brain and language* 96 (2): 221–234.

Pell, M. D. and Kotz, S. A. 2011. "On the time course of vocal emotion recognition." *PLoS One* 6 (11): e27256.

Prinz, J. 2011. "The sensory basis of cognitive phenomenology." In *Cognitive Phenomenology*. Oxford: Oxford University Press: 174–196.

Rakić, T., Steffens, M. C. and Mummendey, A. 2011. "Blinded by the accent! The minor role of looks in ethnic categorization." *Journal of personality and social psychology* 100 (1): 16.

Rakić, T. 2019. "How accent and gender influence perceptions of competence." *Journal of Language and Discrimination* 3: 218–231.

Reiland, I. 2015. "On experiencing meanings." *The Southern Journal of Philosophy* 53 (4): 481–492.

Roessel, J., Schoel, C., Zimmermann, R. and Stahlberg, D. 2019. "Shedding new light on the evaluation of accented speakers: Basic mechanisms behind nonnative listeners' evaluations of nonnative accented job candidates." *Journal of Language and Social Psychology* 38 (1): 3–32.

Scharenborg, O. 2007. "Reaching over the gap: A review of efforts to link human and automatic speech recognition research." *Speech Communication* 49 (5): 336–347.

Scherer, K. R., Banse, R. and Wallbott, H. G. 2001. "Emotion inferences from vocal expression correlate across languages and cultures." *Journal of Cross-cultural psychology* 32 (1): 76–92.

Schiffer, S. 1987. "Intentionality and the Language of Thought." In *Proceedings of the Aristotelian Society* (Vol. 87, No. 1). Oxford: Oxford University Press: 35–56.

Schweinberger, S. R., Kawahara, H., Simpson, A. P., Skuk, V. G. and Zäske, R. 2014. "Speaker perception." *Wiley Interdisciplinary Reviews: Cognitive Science* 5 (1): 15–25.

Siegel, S. 2006. "Which properties are represented in perception." *Perceptual experience* 1: 481–503.

Siegel, S. 2011. *The contents of visual experience*. Oxford: Oxford University Press.

Smith, B. C. 2009. "Speech sounds and the direct meeting of minds." In Nudds and O'Callaghan. *Sounds and perception*. Oxford: Oxford University Press: 183–210.

Smith, B. C. 2010. "Meaning, Context, and How Language can Surprise Us." In *Explicit Communication*. London: Palgrave Macmillan: 92–108.

Speer, S. and Blodgett, A. 2006. "Prosody." In *Handbook of psycholinguistics*. Academic Press: 505–537.

Spreckelmeyer, K. N., Kutas, M., Urbach, T., Altenmüller, E. and Münte, T. F. 2009. "Neural processing of vocal emotion and identity." *Brain and cognition* 69 (1): 121–126.

Wharton, T. 2009. *Pragmatics and non-verbal communication*. Cambridge: Cambridge University Press.

Wilson, D. and Wharton, T. 2006. "Relevance and prosody." *Journal of pragmatics* 38 (10): 1559–1579.

Zhang, H., Chen, X., Chen, S., Li, Y., Chen, C., Long, Q. and Yuan, J. 2018. "Facial expression enhances emotion perception compared to vocal prosody: Behavioral and fMRI studies." *Neuroscience bulletin* 34 (5): 801–815.

Croatian Journal of Philosophy
Vol. XX, No. 60, 2020

Invasive Weeds in Parmenides's Garden

OLGA RAMÍREZ CALLE
University of Saint Louis, Madrid, Spain

The paper attempts to conciliate the important distinction between what-is, or exists, and what-is-not, thereby supporting Russell's existential analysis, with some Meinongian insights. For this purpose, it surveys the varied inhabitants of the realm of 'non-being' and tries to clarify their diverse statuses. The position that results makes it possible to rescue them back in surprising but non-threatening form, leaving our ontology safe from contradiction.

Keywords: Meinong, non-existents, fictions, possibles, impossibles, numbers.

1. *Entering the goddess's forbidden path*

The world of non-existence is surprisingly overpopulated: from fictions to pure non-existents, quasi-existents, and non-actuals, to thick and thin universals, poems and maybe symphonies, maybe numbers, and the whole troop of the impossibles. These 'homeless objects'[1] are all lodged in realms that we know to be in some sense different from the real experiential world, but, somehow, there too. Giving an account of how they can be there in their peculiar, non-existent way is a long-standing puzzle.

To work our way into this ghostly 'jungle' it will be helpful to first distinguish the different sub-species from each other. This can be achieved, for example, by considering the diverse ways in which they relate to what we can characterise as 'ordinary existence' in the actual world. For the sake of drawing the line – which we *ex-ante* assume to be drawable, or else we would not make sense of the problem, nor be discussing it at all – I will specify what I mean by 'ordinary existence'. No entity can be said to exist in the actual world in an ordinary sense if it is not possible

[1] The expression 'Heimatlose Gegenstände' is Meinong's own. See Meinong 1907.

to have some kind of experiential acquaintance with it. This need not be understood, of course, in literal Russellian terms, but it sets indeed a requirement to prove that the experiential conditions of application of corresponding statements are satisfied in order to make them true.[2] Therefore, no entity can be said to exist in the actual world if a) it is not possible to access some spatio-temporal experience that would make claims about it true (if it cannot be said to take place at a given point of time and in a given situation), b) if we cannot regard judgments about it as being actually true on some such basis; c) if it is not possible to re-identify[3] it through different judgements in alternative contexts and d) if it can have no causal impact on other such entities. Correspondingly, if an entity satisfies the opposite requirements to a), b), c) and d), we would agree that it exists. For exposition purposes, though, I will allow talk of 'non-ordinary' ways of existence in describing how entities *are said to be there* other than in the sense described. With these clarifications, let us start our surveillance of non-beings.

a. Fictions (Sherlock Holmes, Pegasus, etc.). These would denote individuals who are invented and taken to exist in invented worlds. They will never exist in the real actual world, nor will their defining properties be instantiated in it. Some real experiential individual might exhibit resemblances to Sherlock Holmes, but he would not be *him*.

b. Pure non-existents (the alien captured by NASA, Vulcano, Phlogiston, etc.). These so denoted guys, unlike Sherlock Holmes or Darth Vader (who were never postulated as potential experiential candidates), were described as existent in the actual world, but we found proof that they do not. Thus, they actually have no existence at all, not even in fiction,[4] nor did they exist and

[2] This requirement, of course, requires some care since non-harmonic concepts (among which are the quasi-existences mentioned later) characterise themselves, precisely, by an imbalance between their introduction and elimination rules that would appear to justify illegitimate existential claims.

[3] The identity requirement, pressed emphatically by both Frege (1884) and Quine (1969), as well as the spatio-temporal condition, that Quine underwrites too, are here necessary though not sufficient conditions, since numbers, for example, would not appear in my account (following criteria a), b) c)) as entities ordinarily taken to exist.

[4] Note that the difference between Phlogiston and a fictional entity is that (unlike our notions of Darth Vader or Madame Bovary) our very idea of Phlogiston considered it to be a chemical substance that was part of the experiential world. That is, it was expected to be experientially detectable (at least at some point and time) and thought to have already a causal impact on the world. Once this is proven wrong, the kind of entity we meant with Phlogiston ceases to exist. Someone might indeed invent a story in which a fictional chemical substance with the name 'Phlogiston' is taken to exist (in the fiction) but, since it would no longer have the pretence to be part of the experiential world, it would not be the same. Such 'pure non-existents' also differ from 'Quasi-existents' in the fact that there is no specific identifiable experience any more on whose behalf they are considered indirectly proven.

pass away. A sub-sort of them, however, about whose existence or non-existence we have no proof (they might already exist or come to existence some time), might be characterised as:

b.1. *Potential existents* (an engineered rational being resulting from the combination of canine and human DNA, a super-intelligent human being with telepathic capacities, unbelievable sorts of super-powerful robots, etc.) could be species of such. Their ambition is to, at some point, have ordinary existence, but they do not.

c. Quasi-existents[5] (individual witches, shamans, saints, etc.). These are peculiar folks, since while they make it possible to identify some real existent beings as their denotation, they mislead us into thinking that what we have identified is a being with properties beyond those which the ordinary existent being actually has. For example, experiencing a person seemingly in trance, claiming to access the spiritual world, acting as a healing figure, using ailments and other non-standard procedures in treating others' illnesses, made it possible to identify a 'Shaman', but on such experiential basis further extraordinary powers were attributed to him.[6] Quasi-existents, thus, never properly existed, but there was a specific experience on whose behalf they were identified. As such an experience it is not merely meant some causal effects in the experiential world that made it possible to presume their existence, as with some theoretical entities; in this sense they differ from pure, exposed, non-existents too. Their relation to ordinary existence can therefore be characterised as misleading.

d. Non-actuals (Socrates; Napoleon; Cervantes; Tolong – the dog I someday will have already been baptised by my niece). These comprise individuals who have been or, predictably, will be, ordinary existent beings, but are actually not such. Their relation to ordinary existence can be characterised as transient.

e. Generalia. Here I will include, in an uncritical way, conceptual meanings and, more generally, the so-called 'universals'. The actual existent world might exemplify them, but they themselves would not have ordinary existence as such and never will. If they exist, they do so in some other realm.

f. Subject-dependents (symphonies, poems, tastes, pains, etc.). Whether these should be considered in this enumeration is much more questionable, but since other authors make a case about them, I will include them too. They would exist in ordinary real-

[5] With this term, I clearly do not refer to what Meinong (1988: 10–11) characterised as 'quasi-beings' (Quasisein). I am, actually, not following any given characterisation at all, but referring to a distinction previously made in Ramírez (2018: 5)

[6] Ramírez (2018, 2012)

ity as much as our bodily feelings in general do; their existence requires our existence, but they are also experientable just the same as all other beings that constitute what we consider standard cases of ordinary existence. They depend upon experiences that we conceptually capture, locally[7] and in a timely manner. This aspect differentiates them from fictions and other non-existents, since the latter are not experienced as such. There is a difference between the thinking process and what the process is about. We can be said to experience thinking but not to experience the entity that Sherlock Holmes is supposed to be. Poems, though, have a mixed character, in the sense that they partially belong to the *fictional* and, to some extent, the *generalia* groups, but full grasp of them requires experiencing their musicality too.

g.Impossibles (the round square, the simultaneously red and green all over book). Clearly, they are not and will never be part of ordinary existence; furthermore, ordinary existence absolutely abhors them. Among all of the non-existent folks, these are by far the most radically threatening ones. This can be for logical or empirical reasons. The above are impossible for logical reasons, but there is an empirical sub-sort:

g.1. *Empirically impossibles* (human beings who do not breath, stars made out of cotton). These are impossible according to what we know about natural laws. Their impossibility to take place in ordinary existence depends on how right we are in our knowledge about natural world laws. They would differ from what I have called 'potential existents' in that, according to our actual knowledge, the latter do not exist, but what we know about the world does not make their coming into being completely impossible.

h. Numbers (0 and the rest). These are as numerous as one wishes, up to infinity, although some of the others could also parallel this capacity. Their relation to ordinary existence is the most intriguing one. Although they do not exist in the experiential world, and never will, they appear to be somehow essential to it too. Unlike *pure non-existents*, they were never postulated as real standard existent individuals, and like *fictions* and *generalia*, their existence, if at all, would be in some other realm.

With this, I think I have a fair taxonomy to start with and to try to make sense of their diverse statuses and of their status as a whole.

[7] Some have argued that music, for example, is a merely temporal experience, but this is disputable since sound waves do seem to be present in some locations.

2. *Meinognian objects or Russellian impostors: Do we have to choose?*

Russell's (1905) analysis of non-denoting expressions in terms of definite descriptions was thought of as a measure to debunk deceiving ontological proliferations of the Meinognian type. This also allowed him to provide a better solution to the problem of negative existential sentences, with which he had earlier struggled. Russell's achievement was, I believe, a huge step towards clarifying when we should be legitimised to assume the existence of what we talk about and when we should not, and how we should conduct a corresponding proof. This is a position abounded upon by Quine (1953) in his critique of Meinong-like views.

Indeed, we can easily be misled by linguistic appearances, assuming the ordinary existence of the presumed referents of our denoting terms. With this, I want to make clear where I stand in this regard, in case what I have to say in what follows should raise any doubt. To expand our ontology, as Meinong (1888) does, to say that with 'the present King of France' we are referring to an 'object' of a peculiar *non-being*[8] *sort*, which *is* already pre-given and is logically prior to our judgement about its existence or non-existence (has 'external Being', *Außersein*: §4) and constitutes some more basic 'grade of Being', is completely uneconomic and confusing. Consider the following quotes

> Should I judge of an object that it is not, so it seems that I have to have apprehended it first to say from it that it doesn't exist (Meinong 1988: 10, §4 'On the Externality (Außersein) of the pure object').
>
> We have to do with a Being, that has neither existence nor subsistence (Bestand), but just to the extent that to both, if one can put it so, as grades of Being, existence and subsistence, a third grade is to be added. This Being would then belong to each object as such. (Meinong 1988: 10).[9]

The idea, therefore, is that in considering, for example, 'the present King of France', we might judge 'that *the present king of France* does not exist'. The that-clause constitutes what he calls an 'Objective', by which we are apprehending an object not merely a representation of it (Meinong 1988: 9)[10] that has Being (in the third grade of our last quote) and which we judge as not existing.

[8] I will use 'Being' and 'Non-Being' capitalized when used as substantives, such as in 'has Being or has Non-Being'. Otherwise I will use the non-capitalized 'being' or 'non-being' both as adjectives in "an object of the non-being sort" or to speak of individual 'beings'.

[9] These are my own translations from the original German text.

[10] That he does not mean merely existence in the representation; see passages such as this:

> (…) that the object possibly needs the representation 'not to exist', just as little as to exists, and even that, just to the extent as it should depend from 'being represented' the obtention at most of an existence—existence in the representation, that is 'pseudo-existence'—as a result (Meinong 2018: §4, 9).

Such an ontology, which embraces all kinds of non-beings, possibles, and impossibles (for which this 'pre-given way of Being as '*Außersein'* holds too), would seem to betray the very sense of the term *ontology* and its allusion to 'what *there is'* as opposed to 'what *is not there*'. Meinong's introduced 'So-sein-objects' (to refer to their having 'a given way of being' even if they do not exist) which he wants to distinguish from 'existent objects' are already more than mere descriptions. They are understood as what the description *is about* and, therefore, as having *Being*. This is, for him, the case with both non-existent objects and possibilities, and also with fictions (such as the golden mountain, which is golden and a mountain even if it does not exist). The difference with '*real* Being' is something which he wants to make palatable by using the contrast between 'So-sein' and 'Sein' – a distinction that, strictly speaking, would work if the former, as the term expresses, is understood as a mere predication, a mere description of a specific way of being. But it rather amounts to seeing the description (which with Russell's procedure would come out as unsaturated) as already concerning an object with a peculiar degraded way of being of the 'non-being sort'. He then can be said to split this elementary Being into the existent and the non-existent (So-sein) types. Thus, we arrive at a conception that to avoid paradox separates the notion of Being from that of Existence, as just seen, and splits then Being into two sorts (existent and non-existent). However, the resource of saying that something *is* (has some objectual pre-given grade of Being) but does not *exist* strikes one as some kind of prestidigitation trick, not far from the claim that 'what-is-not *is*', which already set all alarms by Parmenides centuries ago. Such talk attracted equally harsh critique from Russell, who sharply noticed that even the predication of existence had to be allowed in Meinong's picture of the non-existent objects:

> but the chief objection is that such objects, admittedly, are apt to infringe the law of contradiction. It is contended, for example, that the existent present King of France exists, and also does not exist; that the round square is round, and also not round, etc. But this is intolerable; and if any theory can be found to avoid this result, it is to be preferred. (Russell 1905: 483)

By way of such a reconstruction, Russell attempted to show, in Meinong's own terms, as Bourgeois (1981) argues, the incorrectness of a theory that makes it possible to predicate existence of an object that does not exist and make further contradictory claims. Meinong's attempt to distinguish 'being existent' (as a mere predication) from 'exist' in his reply to Russell (Meinong 1907: §3, 1910: §20) was, in Russell's eyes, just an artificial manoeuvre that disguised one and the same thing (Russell 1907: 439) as two different ones. Notably, even some who, like Bourgeois (1981: 665), tried to make sense of Meinong's response, concluded finding his proposed solution unable to surmount the core of Russell's objection; the objection allowed being reformulated using 'exists' itself as a predication and attributing it to an object that does not exist. Finally, I do not think the nuclear/extranuclear differen-

tiation helps any better to make sense of Meinong's distinction, which Parsons (1974: 574) tried to apply to 'existent golden Mountain'. Nuclear properties are those that according to the description constitute the object, 'being golden', 'being existent', extranuclear ones those that are dependent upon external factors such as *existence* or *non-existence*. The distinction could be said to achieve some clarification of the Meinongian position, maybe, but at the price of acknowledging a predicative mode of existence that ends up being devoid of its very sense, making its attribution worthless. The alternative proposed by Jacquette (1996: 81) to consider existence an extranuclear property and, thus, not to allow its nuclear use, could work better but, as some have pointed out, might require too strong a skimming of the theory for its own liking.

Given such a picture, to conclude that 'the present King of France' is simply a description of ours, for which there is no reference, and no object at all, is a relief. It would unmistakably separate out when what we say is asserted about something that is independently there in an ordinary experiential way, and *just then* has Being and Existence, and when not.

Nevertheless, this convoluted world of Meinong's appears to have something to it too. This explains why a good array of authors, including Chisholm (1973), Parsons (1974), the Routleys (1973), and Jaquette (1982, 1996) or Lambert (1974), to name a few, have defended or logically elaborated consonant views, as well as its recent resurrection in the form of more or less watered down contemporary Neo-Meinongian positions, for example, Priest (2005), Berto (2008), Crane (2012), Eklund (2005). The difficulty is how to place some Meinongian insights so that it is worth all the efforts of his 'ontology' in a helpful way, without making us say what we should not be saying. My purpose, thus, is to try to make sense of this comprehensive picture of beings of the beyond, in such a way that it allows us to keep the gains of Russell´s work. In doing so, I am not going to concentrate directly on Meinong´s proposals, but will proceed rather freely in analysing the different modalities of 'non-existents' above independently.

3. *From fictions to 'reals' and back again*

I will initiate my journey into the realm of Not-Being by considering first 'the fictionals'. I am going to use a little star to distinguish these 'beings'* from those belonging to the realm of Being in some ordinary experiential sense (those that would correspond in Meinong's own ontology to 'objects that have *real Being*'). The * is not meant here to suggest that we have to do with a specific way of being, of the sort that 'is not', as in Meinong, but in the sense that it is not meant literally. Since, as should have become clear, the separation between Being and Existence (or 'there being objects that have no Being') is, in my view, utterly confusing. The peculiarity of such fictional, so-called 'beings'*, is that they actually could not be unmasked through a Russellian anal-

ysis, since they never wore a mask, so to speak. According to Russell´s analysis, the term 'Sherlock Holmes' in sentences such as (1) 'Sherlock Holmes played the violin', is to be seen as a definite description (recording the main features attributed to Sherlock Holmes in Doyle's novels). We would then prove that there is no such experiential individual that satisfies it. Without entering the difficulties that such a descriptive account of proper names is seen to have in the later discussion,[11] I would say that the real issue here is a different one. Let us reconstruct the sentence above in Russellian terms, where SH stands for Sherlock Homes and V for 'plays the violin'.

$$(1) \quad \exists(x)\big(SH(x) \wedge V(x)\big) \wedge \forall y\,\big(\big(SH(y) \wedge \mathrm{V(y)}\big) \leftrightarrow (x = y)\big)$$

The second part of the sentence expresses the uniqueness and identity of the so-described 'being'*. What is the problem then? In the non-fictional case of 'the present King of France', or, for a change, 'the first female President of Russia', a Russellian reconstruction makes clear that there is no experiential individual satisfying the description and therefore no such real existent being we are talking about. However, it might be that these sentences uttered in some other moment of history come out true. It is possible that we might, on some occasion, find an individual who corresponds to the description. So, here we take the sentence 'the first female President of Russia has a dog', where the FFPR stands for the subject and D for the predicate.

$$(2) \quad \exists(x)(FFPR(x) \wedge D(x) \wedge \forall(y)((FFPR(y) \wedge D(y)) \leftrightarrow (x = y))$$

We expect this sentence to be satisfied by an existent actual being that exhibits those properties if we are to find the sentence true. Comparing this now with the sentence (1) above, we see that although it has the same form, the peculiarity of (1) is not that it is not satisfied by any existent individual having the characteristics attributed to the assumed Sherlock, but, rather, that it could not be. Why? Not because there could be many, as Kripke once argued (1980: 58), but because, as the second part of the sentence makes clear, there is only one such being, and even if we should find another satisfying most of the characteristics of Holmes, it could not be *him*. It is a characteristic of the authentic and unique Sherlock Holmes that he is fictional, and an ordinary existent one would not have such a character. Therefore, while sentence (1) does presume the existence of an individual we are talking about, it is not open for saturation. It is rather already satisfied by our fictional

[11] Actually, I am not interested at present in the difficulties of determining which description would be the adequate reconstruction of a proper name, or how it is that individuals with differing information manage to refer to the same individual. Rather, my focus is on how we prove whether the individual who we are attempting to refer to by some, call it, 'ideal description' is there or not. Of course, there can be differences in the particular aspects being proved by certain individuals depending on the information they have. Nevertheless, the way they are proved, in general, by human beings, can be the same. This is so even in the limited case where all someone might have as a description is that there was someone baptised as 'Holmes'.

individual. This means that in speaking about Sherlock Holmes, we do not have a normal unsaturated sentence but an already saturated one. In contrast, sentence (2) is and should remain unsaturated by any positive object, as long as no existential individual satisfies it. However, if we reconstruct our sentence (1) in the following way, including F as a predicate for fictionality,

(1) * $\exists(x)\big(F(x) \wedge SH(x) \wedge V(x)\big) \wedge \forall y\,(\big(F(y) \wedge SH(y) \wedge \mathrm{V}(\mathrm{y})\big) \leftrightarrow (x = y))$

and then we take ourselves to have a fictional being named 'Sherlock Holmes' satisfying the sentence, we fall back into the contradictory claim that *there is* (or exists) a *being* (without star, since we are talking about existence without qualification) such that he has the extra property of being fictitious. Then, since 'fictitious' means 'not really existent', or so we take it to mean, we obtain our paradox. We actually go directly to what Russell found 'intolerable' in Meinong's view of the idea of 'objects that do not have existence', objects that are graspable, and in a way are externally there but do not exist.[12] But if we reconstruct the sentence as in

(1)** $^{\mathrm{Fic}}\,Ex\big((SH(x) \wedge V(x)\big) \wedge \forall y\,\big((SH(y) \wedge \mathrm{V}(\mathrm{y})\big) \leftrightarrow (x = y)))$

using a fictional operator to apply it to the existential quantifier and all that follows, then both the description and its saturation by our *presumed* existent unique individual come out as fictitious. Being fictional is not here predicated of a pre-given object, but through the operator *Fic* it is meant to characterise existence itself and all that is attributed to it (the objects that are part of existence, their properties, and their relations). Existence is not a property but a presumption and, therefore, merely fictional. Furthermore, we should separate out here too the existential *claim* (the existentially quantified sentence) and the existential *presumption* that there would really *be** a unique object that saturates it: the presumption of an individual, call it $^{\mathrm{Fic}}a$, that does satisfy (1)**. That is, we need not just *say* that such and such exists, but to represent it to some extent as an existent individual who would, for example, occupy a different spatio-temporal point than the other individuals in a fictional play, and picture his continuity and spatial moves along the story, his interactions with others, and so on. The proposal, therefore, departs from the nuclear/extranuclear distinction or the encoding/exemplifying view, since it does not amount to acknowledging a merely predicative mode of existence.

[12] Such a formulation might suit the case where an existential claim is made that describes an apple with such and such properties, and we then saturate the empty space with a fake apple as the unique one satisfying the specific characteristics of the description. Such an 'apple' could be seen as really existent though as a fake. It might be supposed that regarding a painted picture of Sherlock, put in place of the description could do for its existence as well, we could pretend it is the individual we are referring to. But since the paper figure would not satisfy the descriptions attributed to Sherlock Holmes, it would not do either.

To some extent, the reconstruction presented in (1)** might conform more likely with certain forms of pretence accounts (such as Walton 1990). But, for now, I would not want to commit myself to any label until the complete contours and implications of the position I shall be defending are in place. The specific formulation chosen in (1)**, regardless of possible coincidences with other proposals too,[13] is thought to serve the interests of the view I am leading to and should be read in the terms proposed.

Interesting here, from my perspective, is the perfect analogy of the structure of fictional discourse, whose definite descriptions are already satisfied by a unique individual, with respect to the standard practice where ordinary existent individuals are assigned to corresponding existential sentences. What makes Sherlock Holmes more than a mere description is precisely the (fictional) presumption of a unique individual who would exist and who satisfies the description. Mere possibilities (*potential existents*), such as 'the individual who is a combination of human and canine DNA' or 'the golden pencil on my table', are not presumed to have any individual satisfying corresponding descriptions. As we put it at the outset, 'possibilities' aspire to ordinary existential satisfaction; correspondingly, their existential sentences would only be fulfilled if, for example, a real golden pencil is to be found on my desk, turning the possibility into a reality and the sentence into a true one. If we, therefore, subject them to proof in the Russellian experiential sense, we find corresponding descriptions to be empty. It is thus important not to attribute non-saturated definite descriptions positive denotation (contrary to what Meinong does with his So-seins in direct predications) and to trace a clear distinction between such merely possible statements and fictions. This distances my proposal too, I believe, from Meinongian-based free logic accounts (Lambert 1960), since I do want to keep differences regarding commitments to existence.

Now, the case of 'non-actuals' brings particularly interesting aspects to the problem. While it is easier to assimilate future beings to mere possibilities and understand them along the same lines, past individuals raise interesting questions. It seems to me very weird to think of my deceased father as a mere description. I actually consider myself as being able to refer back to *him*, to refer to the individual who was there. But how can that be? Kripkean 'causal chains' tell a story of me referring back to the individual who was baptised 'XYZ', and whom I call 'my father'. But the above does not say much about how this is possible, nor does it amount to any commitment regarding the acquired status of the referred object. Actually, if we look at it, the whole past itself does not exist at all! For all that is said, he could perfectly well be counted as a Meinognian object with non-existence. If, on the other hand, I use Russell's procedure, a corresponding existential sentence would come out false. There is (unfortunately) no such person to be encountered anymore. But now what has suddenly struck me as strange

[13] For example, other defenders of the pretence view, such as Recanati (2000).

is how I have been referring to him all along… as he was there. What has really changed? First, there were times when I was away for a long period and could not continuously check on him. He could have not even been there, and I would have kept referring to him just the same. My knowledge of his *past* existence is no warranty that I am referring to an existent being. More worryingly still, the problem expands when considering that we have in fact no warranty at all; neither can we be said to be referring to an existent being if all we have is the past. The whole issue ignites then when thinking that this does not just happen with my father, but with all and everything else. How do I keep referring to it all, if my presumed existent beings are not literally existent for me after they pass my surveillance – my experience of them? Testimony can occasionally serve this purpose, but I do not keep reconfirming through others the persistence in Being of my objects of reference either. What about whole cities and oceans and my very past infancy with its populated ingredients, places, and encounters? This is a real black hole in our tidy permanent picture of world stability – one we will inescapably have to deal with.

On the other hand, while sentences about 'the golden pencil on my desk' would be satisfied by finding a unique golden pencil on my desk, no future found individual could satisfy the description of my father. He is a *complete* individual (in Meinong's terms)[14] in a way 'the golden pencil on my desk' is not (despite the unique character of the definite description). He is also complete in a way Sherlock Holmes isn't. No new individual could, therefore, take his place. In this sense, non-actual past beings clearly differ from mere possibles and fictions; of whose 'existence' we merely now as much as the description states. But this uniqueness and the once-and-for-all saturated character of descriptions of my father by him do not solve the question of non-actuality, which, as we just saw, expands also to ordinary existent beings of which we just have past experience. Actually, if we consider this problem in its threatening radicality, there appears to be no way to give an account of the existential stability of our referents, the stability we experience them as having for us, by attributing ourselves a capacity to 'keep them present'. To keep, actually, our whole world present beyond the continuously fluent and transient character of our experiences. That is, to maintain something like an imaginative representation of our world while it decays — one where the different real interacting objects are attributed, at a representational level, different non-colliding positions in space and are expected to have a continuity through time and mobility consistent with that of our own. If this is how we experience it, we must be able to situate these objects and their moves and relations as if in some sort of land cart.

[14] According to Meinong, while real individuals are complete, fictional individuals and possible ones are not. We just have as complete a characterization as the description goes and that leaves many informational gaps: did Sherlock have an ant? Was he ever in Albacete? There can be no true or false answer to that.

This stable 'representation' that carries its complete non-actual referential objects as something more than mere descriptions[15] (and in that literal sense I mean the word) would very much resemble the kind of representation we reconstructed for the fictional case, where it was already suggested that making sense of fictional claims implies already a similar figuring capacity. This is especially striking in today's virtual games, where we interestingly develop exactly that diagramming in a visually 3D representational way through which we put down our own way of conceiving it. However, maybe we should rather consider that if we can come to such a representational picture in fiction, it is because we already have at our disposal those representational capacities, since we need them to constitute what we consider our referential 'reality'. Put differently, that reality, our reality, is fiction! This means that what we would be referring to when speaking of an existent being such as that referred to by 'the ex-President Obama' in the sentence 'the ex-President Obama is married' should be reconstructed along similar lines to (1) **. With XPO standing for ex-President Obama and M for married, we would obtain:

$$(3)\ ^{**}\quad {}^{\text{Fic}}\exists(x)\big(XPO(x)\wedge M(x)\big)\wedge\ \forall y\ \big(\big(XPO(y)\wedge \mathrm{M}(y)\big)\leftrightarrow(x=y)\big)$$

where existence is postulated again at a representational level and presumed. Presumed too is the existence of a corresponding object satisfying the statement, as ex-President Obama himself is. Of course, I cannot literally mean that there is no difference between reality and fiction, and I do not. What I am saying is that referring to reals requires such a representation too (in the alluded literal sense of representation, beyond the mere putting into words or, counterpart, thoughts), though not alone. Therefore, since I am proposing a common reconstruction for reals and for fictions, I will change the fictional modifier *Fic* of the existential sentence above into a presumption modifier *Press*. Corre-

[15] I found Recanati's (2000) proposal to solve this problem by appealing to descriptive files very interesting, and he is probably right that we store files that way too. However, I think his proposal is not enough for several reasons. First, such files need not imply taking for existent and referring all along to the individual the file is about, the one satisfying the corresponding descriptions. Second, referring to such an individual as existent requires us making sense of its position and its possible further moves, possible future encounters with ourselves or others in other places in a consistent and timely manner. This requires, I believe, activating some spatio-temporal representational coordinates. Therefore, this position might be complementary to his. Actually, Recanati (2018), in discussing parafictional discourse, poses precisely the problem I allude to in defence of a 'presumption' instead of a 'pretence' view. The parafictional use possibly allows a better appreciation of the point. As Recanati stated, in parafictional uses, we are not pretending; we are taking the claims to be true about a given object, for whose purpose the representational object in my picture very well serves. Of course, since we move at the representational level, neither objects nor properties are real, but just in representational modus. Actually, Recanati (2000) mentioned a proposal that defends the existence of 'dot objects.' I do not know enough about that position, but maybe it comes closer to what I am proposing.

spondingly, (3) ** turns into $(3) \ast\ast\ast^{Press} \exists x\,(XPO(x)\ldots)$. Moreover, (1) ** turns into $(1) \ast\ast\ast^{Press} \exists x\,(SH(x)\ldots)$. The notion of 'presumption', unlike, for example, 'pretence', implies not the faking of something, but the taking of it for granted, which I think is not the same.[16] Actually, I think that in his talk of the 'parafictional' Recanati (2000) would be acknowledging this difference.[17] According to Recanati, in parafictional uses of language we are not pretending but rather taking for true. This is akin to what I mean with 'presumption'. We presume existence at the representational level. We presume that these beings exist throughout time, and we keep them represented as doing so, while this would just be (most of the time) a mere presumption. This is also the case in fiction, since our attitude in fiction is not that of mere supposing, but the attitude of something being the case, being true. Therefore, it is important to specify that with the term 'presumption' I am implying the idea of treating the corresponding individuals as real, postulating it to be the case, as opposed to merely supposing or entertaining a possibility that we do not treat as true. But, if 'ordinary existence' is to be seen as a presumption in this sense, then how do we tell the difference between fictions and ordinary existents?

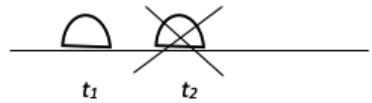
Well, we should ask Russell. As he says, this is a matter of acquaintance, of the origin of the representation being based in direct acquaintance. Not understood literally in the sense Russell meant the notion of acquaintance, as we said, but just the idea that it is a form of experiential encounter that gives us grounds to claim that there is a correlation between our stable representation of the individual with something out there. We have, therefore, to do with a *well-grounded* presumption, which is what we call 'real existence'. That is, the difference between the fictional and the real case is not at the level of the type of representation but is due to the grounding, or lack of it, of the representation. It is the mixed character of the real, perfectly experiential, on the one hand, and representational, on the other, that makes the difference.

So, if the representation, in (6) below, goes for both fiction and real existence, where our picture is kept stable throughout a lifespan (until we are told or experience otherwise), real existence departs from a prior time-fugitive originating grounding layer (5).

[16] In my account we are not simply linguistically pretending that there is an individual such and so, and pretending sentences about it to be true, but we are actually literally representing some such individual occupying a given space, having a continuity through time, etc. and truly referring to it.

[17] See footnote above on Recanati.

(5) $t_{fugitive}$



(6) t_{stable}

At the experiential level in (5) at t_2, our object is not there for us anymore, but at the representational level in (6) it remains in representational form as our referential object. The time-limited real experience of *ex-President Obama* builds the well-grounded claim that our stable representation of him is real, as long as we (or someone else, in which case knowledge of direct experience is obtained by testimony) have a new experience to the contrary. Therefore, if we subject our claim to the Russellian proof at the experiential level:

$$(3) \quad \exists(x)\big(XPO(x) \wedge M(x)\big) \wedge \ \forall y \ \big(\big(XPO(y) \wedge \mathrm{M}(y)\big) \leftrightarrow (x = y)\big)$$

It comes out positive as soon as we can provide an acquaintance experience or well-grounded testimony of such. But the stable further existence of our referent, when considering it in the modified form of (3)***, allows confirmation in the presumed modus too. Therefore, a claim of existence on the basis of such a stable representation, while it is a presumption, is a well-grounded one, since at some points of time we did have existential satisfaction in the Russellian sense. To put it differently, at the level of experience, where we situate ourselves, the very existential aspect of our claim in (3)*** $^{\text{Press}}\exists xists$ finds sufficient confirmation in (3) $\exists xists$,[18] when satisfied by an individual *a*, while the satisfied existential claim in (1)*** does not have a parallel confirmation in (1). Therefore, our representation in (1)*** is known to us from the very beginning, not to be *of* the experiential reality. It is just our representational attitude that is of the same kind. In a sense, it is as

[18] This mixed character explains why *at the experiential level* in the moment of acquaintance we can directly refer a lá Kripke to 'Gödel', for example, while later on any explanation of who we mean, at that level, for those who lack direct acquaintance, must be in descriptive terms, even if the description reduces to the individual that was named Gödel at t_n in place p. At the same time, though, we have a sense that we can keep on directly referring to *him*, because at the representational level, in keeping a corresponding representative of Gödel, we do.

if our $^{\text{Press}}\exists xists\ quantifier$ acquires the status of a place holder, that is of a higher order *'quantifier-variable'* of existence itself, that can be saturated or not by the real $\exists xists\ quantifier$.

On the other hand, the type of justification for our claim in (3)*** is not just based on encounters, although it primarily is, but also a matter of knowledge. The knowledge is that unless there is a sudden decease, living beings continue to exist as we ourselves do. I am absolutely not pretending, though, that we would be doing all this labour of deduction consciously. It is more a necessary condition of possibility to be able to refer in a non-discontinuous way and also plan action and interactions with others towards the future. But in becoming aware of this, we can explain that we are well justified in doing so, keeping the representational stability of existence through the procedures mentioned. In some cases, where our experiential confirmation is really a continuous presential one, paved through continuous encounters and reconfirmations of existence, the presumption is almost equal to experiential reality, as happens with our own self, whose past we keep stable too and whose present existence we continuously experience.

In the case of deceased non-actual beings, what we become is proof of their ceased existence. What we have is a (well-grounded) presumption of existence whose continuity is truncated up to a given point. Therefore, representational reference to the past, since stability makes it possible to refer back and forth to an individual, is still of the real *existent* kind, as it alludes to a time when there was still the grounded level of proof. But nothing of the sort is possible anymore in a present or future form.

Notice that the proposal here is an epistemological one. It does not limit itself to say in the case of real existence, that there is an individual out there we take our semantics to refer to, as some authors have argued (for example, Parsons 1974). The idea is more radical in the sense that of necessity, the real existent object is just stable as a such for us through a representation. This means that we are not referring to an existent individual in the experiential world all along and then we give a semantic account of it. But, rather, that our experience merely delivers a punctual basis and the representation is required for stability.

4. *Numbers*

As Frege (1884) well saw, counting unities requires separating them out through a concept.[19] This, I believe, need not mean literally that we need 'linguistic concepts' to distinguish unities, as in some direct realist positions. [20] It also allows for a more basic pre-conceptual reading

[19] 'Denn der Begriff, dem die Zahl beigelegt wird, grenzt im Allgemeinen das unter ihn Fallende in bestimmter Weise ab. Der Begriff "Buchstabe des Wortes Zahl" grenzt das Z gegen das a, dieses gegen das h, u. s. w. ab.' (Frege 1884: § 54, 42).

[20] To this point too Ramírez (2020a: 160, footnote 27)

according to which there is no counting that is not done from a given perspective, by distinguishing some particular aspect on whose behalf individuals can be separated out from their surroundings. However, for the sake of simplicity, I will be referring to the 'conceptual' sorting out of unities. Yet, if we are to consciously become aware of what it is that we do in this process, we would be representing to ourselves this very operation (the conceptual detachment of a unity) in general terms, a procedure through which we figuratively capture our own activity. That is, we do not cognitively appraise it, as if it would be some further occurrence or experienceable phenomenon in our epistemic world, but rather we reconstruct what (we think) must be taking place through our own performance. In doing so, two aspects should be distinguished: the very conceptual individuation and the real individual that results. Now, as I have argued elsewhere (Ramírez 2020a:160, 168), I believe that, contrary to what Frege ended up capturing, it is the latter that constitutes the proper numerical unity. It is not the formal conceptual aspect, something that certainly would be equivalent to a class, but the corresponding representation of a real individual or a real object cut off from an extended background. To put it differently, it is the very stable representation of a (non-further saturable) real individual that constitutes the numerical unity. It is a representation, but a representation of a reality. This is precisely, I think, the same kind of individual, of a singled-out object, we refer to, in both fiction and our representation of reality – not concepts, not classes, not mere descriptions, but the representation of the *reality* that satisfies our quantoren, and that is exactly the same as the numerical unity. While Frege, I believe, was initially orientated towards this, he shied away, since this would have brought him beyond the mere formal, analytical account to a representational one (more akin to the Kantian representation in Intuition).[21] He miscalculated in thinking that the concept of the unity itself (or of a class of them) could be a substitution instance of further existential sentences, with the known disappointing consequences.[22]

However, what results out of the considerations made, to put it concisely, is that Sherlock Holmes, after all, is just…a number – nothing we should wonder about, since he always was a master of disguise. But we should not feel too superior either, since, at the end of the day, we are a number too! But one of a mixed sort: half ghost, half real. Nevertheless, there are still 'others' in our list that might have it worse.

[21] I use 'Intuition' with a capital I to refer to the Kantian notion, to separate it out from the mere idea of 'intuition' as a presumed alternative way of acquiring knowledge.

[22] Ramírez 2020a: 161–168.

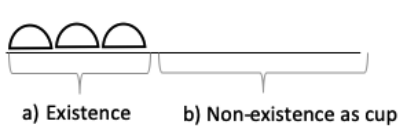
5. *Pure non-existents*

What are we to say of the family of 'Vulcano' and the like? We cannot presume their existence, since what we are presuming is that they have none; not even fictional. Giving an account of their predicated 'non-existence' has always been especially problematic. In Meinong's ontology, we are saying of an 'external object', Vulcano, that non-existence is attributable to it. We obtain, therefore, the contradiction that *there is* something, we are denoting some object, that *is not*. Russell's alternative reconstruction dispelled the puzzle, showing that it was simply a non-satisfied description. So, we have an existential quantified sentence that proves to be non-saturated at the experiential level; just as in the case of Jupiter, we have one that is. A reconstruction of such sentences as if they would nevertheless be denoting some (low grade Being) objects, while posing no problem in fiction and by the stable representation of real existent objects when they are not present, is here to be avoided. It would block their capability of being saturated at the experiential level, which is the only one that could satisfy the kind of existential claim made through a statement postulating the existence of a corresponding planet. According to our knowledge at such a level, the sentence would be empty and, if we consider our proof of non-existence to be definite, then the very possibility considered would be denied. If this is the case, then we should conclude that: it is not possible that there is the object such and so.

But what happens when we have no proof one way or the other? What happens when we merely *entertain* the possible existence of an entity such as 'the nearest by planet containing other intelligent beings'? These are finally the 'possible beings' we have alluded to but not yet directly treated. As opposed to real existence and fiction, we are not representationally taking for granted their actual reality. That is not what we want to presume. We *know* there are no well-grounded reasons to claim real existence (nor to claim the contrary), and as opposed to fiction, we are not presuming them to be objects at the representational level either. Nevertheless, here too, we want to have the opportunity to consider them at the representational level, we actually do, in order to be able to talk about and operate with them, to consider future interactions with others or further consequences that their existence, actions, or reactions could have for us in our world. This is like if we would keep a chess game board and consider moves back and forth to analyse what their consequences would be. In considering their *possible* real existence, we are entertaining not their actual reality but some possible alternative or future one. But how are we to do this without providing such sentences with a denotation, which, if at all, they ought to have (corresponding to the claim made) at the level of experiential reality? We cannot merely represent such an object as a presumption of existence, well-grounded or not, because that would confound it with real existents or fictions. Therefore, at the representational level, this must be reconstructed differently.

That is, at the representational level, we pictured the resulting unities of conceptual individuation (those of saturated existential Russellian sentences) as being what we assume to be stable (well-grounded) realities. These are taken to have a corresponding experiential grounded level. Fictions, we said, differ in terms of their lack of this mixed character and grounding source. They are mere representations of existence. But we do just the same as we would if they were there, it is just that we are aware of their merely presumed character. This cannot be equally reconstructed in the 'entertaining future or alternative reality' cases, since we are in no way representing actual existence.[23] We are, in fact, representing lack of existence, *lack of object*, since possibilities define themselves by 'not being taken to be the case'.[24] But how can that be? Here we have our puzzle back. This was, I guess, the difficulty with which Meinong was confronted. Although he saw in this sense no essential difference between fictions and possibilities, the quandary was to picture something that-is-not as if it would be. Yet, before, when giving an account of what it is that we represent when we depict an individual as existent, I argued that, ultimately, we had to do with numerical unities. If we say there are three cups on the shelf, we distinguish each cup through the concept and obtain three unities of the sort. What remains outside conceptual individuation has no existence 'as cup', so it is nothing from the perspective of the concept.[25] If we picture the existent unities of the cups, like in (4) a), below, and if we want to entertain possible cups that do not exist, this would amount to representing them in the territory of non-existence in (b), but relative to the concept of 'cup'.

(4)



[23] Unfortunately, I cannot enter to comment on the differences of my account here with the modal proposals offered by Berto (2008) and Priest (2005), but I guess they will be apparent.

[24] We can distinguish epistemological and ontological uses of 'possibility'. This can be due to our knowledge: something possibly exists, but we do not know. Or it can be an ontological matter: something does not exist, but it possibly could. The claim affects surely the second case. However, since we are the ones who represent reality according to our knowledge, at the practical level ignorance of whether something is the case amounts to not counting it as real either.

[25] I am basing myself in this section in both, Ramírez 2020a and 2020b, here specifically in 2020a: 169–170).

(4)*

a) existence b) non-existence as cup

So, if we entertain the possibility of there being two cups on the shelf, for example, that means presuming the *non-existence* of corresponding objects (to remain consistent with possibilities not being existent).[26] However, in the sense that *there are no objects* of the sort and, thus, being consistent with Russell's claim that such denoting expressions are not saturated, what we are doing is representing them like in (4)* above, as non-existent objects but as cups. Notice that there is a difference between saying: (i) 'there is an object that does not exist', and (ii) 'an object does not exist'. We are saying (ii) and in this sense we talk of non-existent objects or non-existent cups, while Meinong was seduced to talk about (i). Now, if, as I stated previously, the general representation of numerical unities amounts to the representation of the application of a concept to some reality, what are these ghostly unities of non-existence? Well, I have argued (Ramirez 2020a, 2020b) that they are the negative numbers. Negative numbers are absences: minus 1, minus 2, minus 3 objects, which might be cups or whatever, if what we are considering is the non-existence but possibility of three cups on the shelf, for example. This mere representing of *the lack of cups* is interesting from the perspective of searching for their reality. After all, it is just through a concept, as Frege stated, that we can distinguish unities, and it is through the concept of 'cup' that we are making 'wholes for cups in the emptiness' so to speak, that are nevertheless not to be counted 'as cups' and can be counted as cups not being there, and thus negative cups. In this way we can accommodate Meinong's notion that the cups are 'so-seins', that is, they have the predicative form of 'being cups' and, as he argued, 'you can count what doesn´t exist' (1988: 5). There are two negative numbers of cups on the shelf. If we assume this is right and go back to our Vulcano case, what is the difference between Vulcano of whose non-existence we have proof and the possibilities just considered?

If we pay attention, the difference between both parallels with the distinction before between fictions and realities. At the representational level, they are the same, but at the grounding level, we have a difference. Correspondingly, in the Vulcano case, we would be representing it just the same as by possible negative existents, with the difference that

[26] Compare with Ramirez (2020b: 18–19)

here we have a well-grounded proof of non-existence, the non-saturated Russellian sentence at the experiential level, in addition to proof of its non-existence. We can thereby make sense of the 'negative number' Vulcano, which we can say does not have real existence on the basis of a proof according to the Russellian method. The mixed character of the representation – half experiential (proof of absence at the experiential level), half ghost (negative number) so to speak – applies here too. However, if the proof is definitive, as we stated previously, we can simply deny $Not - \exists xistent^{Press}$, that is, the negative existential presupposition as a whole.

I am not going to enter much into the realm of the quasi-reals, since actually an explicit Russellian sentence at the experiential level would exhibit the purported satisfier as a hoax. Therefore, the status of 'quasi-real' would thereby have to be abandoned, and we could turn the corresponding individuals into fictions, attribute them a non-literal sense, or make them disappear. Although understanding how such terms come into being and their deceiving force is quite interesting, in our context they do not add much new. The case of *generalia*, however, I will leave for a future occasion, since although they would behave as reals in being partially well grounded, they pose a particularity of their own.

A different issue is with the impossible, which raises much perplexity. As opposed to pure non-existents, with impossibles it is not that we have proof of their non-existence. It is that we need not even try, since we know in advance that no proof could be found. One option would be to see impossibles as irrational numbers, which do express the impossibility of a numerical distribution in any conclusive numerical terms, obtaining a number that, in some sense, is not a countable number at the same time. It is a magnitude not capturable by one unity or the other, not expressively by any discrete counting, since whenever one splits a continuum, one originates a difference between the unities and the continuum. This is often irrelevant, but sometimes it is not.

So, if an extension is a continuum between two discrete unities – call them 'a' and 'b' (which can also be understood as parts and parts of parts of unities) – there is always a difference between the numerical division and the 'not-numerically divided', and therefore something in between the split parts. Since it has no presence as any kind of countable unity that can be considered 'something' (it is no expressible number), it does not have existence as either the one or the other; or, alternatively, it could be included in either. In a sense, it can be considered both 'a' and 'b', but paradoxically 'a' defines itself by 'not being b', from which it has been detached. So, it could not be 'b' but, at the same time, it is not something 'other than b', and so it can also be 'b'. But, then again, 'b' defines itself by not being 'a', and so it cannot be 'a'. However, since it is not 'something other than a', it can also be 'a'. So, we obtain this contradictory result that it is 'a' and 'b', but if it is the one it could

not be the other. Just the same as the 'round square', we figure out something in the middle of the opposition between round and square – something that actually cannot be, since if it is round it is not square, and vice versa. Even at a predicative level, and even as a 'so-sein' in Meinongnian terms, it is contradictory, since it is *so* and *not so* at the same time. Is it not precisely this position we are somehow imagining? Could that do? I will leave it here, since I am aware this might strike one as slightly strange but do please give it thought.

With this, I believe we have ordered our ontological picture without doing wrong to the extremely important distinction between the existent and the non-existent in Russellian experiential terms, while at the same time making room for some of the Meinongian insights. Meinong's non-existents (including fictions, possibles, pure non-existents, and impossibles) do not have room in our ontology, for which experientiality, as both Russell and Kant claimed, is the basic requirement. Their presumed existential quantor is a non-saturated one, as we might put it. Unlike Meinong, who explicitly denied we had to do with representations (1988: §4, p. 9),[27] we assigned them a place at the representational level, where picturing what is not part of the experiential world and distinguishing it from what is, is not absurd; furthermore, we came to conclude that even the stability of our actual referential world owes much to such a representational level too. Reality, as we understand it, would not be the reality it is without it.

6. *Final thoughts*

The picture that emerges from the deliberations above is one in which our representational thinking about objects – real or fictive, possible, impossible, or non-existent – which both Meinong and Frege (with the exception of the impossible in the last case) took to be objective, shares the same basic structure as our dealing with mathematical numbers. That is, in thinking about some such individual, we would be applying a certain formal structure to some units that are made available in representational modus as something external, and we are ultimately operating with numbers.[28] In addition, the other way around, when we do mathematics, we are, at the end of the day, thinking about reality and its possibilities, or even impossibilities, to better understand it.

Pretending that with our rational reconstruction of such a representational level (which we take for granted but do not experience as such epistemically) we would have to do with an enhancement of the world of experience amounts to the same kind of failure Kant so carefully separated away in his *Critique of Pure Reason*. The confusion is between transcendental reconstruction and transcendent knowledge

[27] Text quoted in footnote 10.

[28] I would not want to commit to fictionalism in mathematics, since this representational aspect is not simply pretence, but belongs to the very constitutive way of our knowledge of the world.

beyond the realm of experience. In the first case, we are not acquiring any real knowledge. We are rather reconstructing what has to be the case for this cognitive activity, characteristic of our knowledge and thinking about the world, to be the way we experience it. Moreover, we do this in the cognitive terms we understand, figuring out the conditions that would make it possible, just as we try to figure out the likely forces behind some events in the world. Of course, we ourselves belong to that experiential world, as does our brain, but we cannot properly acquire experiential knowledge about how we think. Thinking is an experience, but how it occurs is not,[29] just as are experiences those of other internal events, such as pains, pressures, or emotions, not our capturing of them. We can have second-order thoughts about our speaking and thinking or our emotions of which we are aware, as some have suggested,[30] but there is no experience of how we come to do so. This is more a matter of inferring that we must have done it, since we are aware of them. The representational structure of our thinking about existent and non-existent objects or numbers is something we would become aware of in an inferential sense: it must be like that. We know we think in such a way, but how we manage to do so is a later re-enactment. If we are thereby right, it does take place, but our knowledge of it is no direct experience, and we cannot claim to *know* it to be. Therefore, it is not true that there are any such external objects or numbers, the weeds ought to be purged in this sense, not merely domestically arranged. But it is true that we must represent them as being (or as lacking).

Finally, it might be claimed that our own representational forces are part of the world, and they indiscriminately are so, just as any other forces we exert, but they generate no further realities beyond themselves.

References

Berto, F. 2008. "Modal Meinongianism for Fictional Objects." *Metaphysica* 9: 205–218.

Bourgeois, V.W. 1981. "Beyond Russell and Meinong." *Canadian Journal of Philosophy* 11 (4): 653—666.

Chisholm, R. 1973. "Homeless Objects." *Revue Internationale de Philosophie* 27: 207—223.

Crane, T. 2013. *The Objects of Thought*. Oxford: Oxford University Press.

Frege, G. 1884. *Grundlangen der Arithmethik*. Hamburg: Felix Meiner Verlag.

Frege, G. 1884. *Foundations of Arithmetic*. 2nd ed. New York: Harper and

[29] A similar line of reasoning concerning this point is to be found in Ramírez (2020a: 176–177)

[30] In a sense, Recanati (2000) captures this idea through his notion of 'metarepresentation'. However, I am not completely sure if he understands it the same way. To be sure, the notion of representation in this paper, more akin to the Kantian representation in Intuition, is different from that in Recanati's picture.

Brothers Harper Torchwoods (1953).
Jaquette, D. 1982. "Meinong's Theory of Defective Objects." *Grazer Philosophische Studien* 15: 1—19.
Jaquette, D. 1996. *Meinongian Logic. The Semantics of Existence and Nonexistence*. Perspectives in Analytical Philosophy 11: Berlin–New York: de Gruyter.
Kant, I. 1781. *Critique of Pure Reason*. Cambridge: The Cambridge Ed. of the Works of E. Kant (1998).
Kaplan, D. 2005. "Reading 'On Denoting' on its Centenary." *Mind* 114 (456): 933—1003.
Kripke, S. 1980. *Naming and Necessity*. Cambridge: Harvard University Press.
Lambert, K.1974. "*Impossible objects.*" *Inquiry* 17: 303–14.
Lambert, K. 1960. "The Definition of E! in Free Logic." *Abstracts: The International Congress for Logic, Methodology and Philosophy of Science*. Stanford: Stanford University Press.
Meinong, A. 1888. *Über Gegenstandstheorie Selbstdarstellung*. Hamburg: Felix Meiner Verlag.
Parsons, P. 1974. "A Prolegomena to Meinongian Semantics." *Journal of Philosophy* 71(16): 561—580.
Priest, G., 2005. *Towards Non-Being. The Logic and Metaphysics of Intentionality*. Oxford: Clarendon.
Quine, W. v. O. 1953. "On What there is." In his *From a Logical Point of View*: 1—20. Harvard: Harvard University Press.
Ramírez, O. 2012. "Boghossian's Blind-Reasoning Conditionalization and Thick Ethical Concepts." *Ethics in Progress* 3(1): 31—52.
Ramírez, O. 2018. "Between Witches, Angels and Unicorns. The Possibilities of Expanding Russell's Existential Analysis." *E- Logos* 25 (1): 4–15.
Ramírez, O. 2020a. "Numbers, Empiricism and the A Priori." *Logos and Episteme* 11 (2): 149—177.
Ramírez, O. 2020b. "I do not believe in Meigas but there are such." *E-Logos* 27 (1): 4–20.
Recanati, F. 2000. *Oratio Obliqua, Oratio Recta*. Cambridge: MIT Press.
Recanati, F. 2018 "Fictional, Metafictional, Parafictional." *Proceedings of the Aristotelian Society* 128 (1): 25—54.
Routley R. and V. 1973. "Rehabilitating Meinong's Theory of Objects." *Revue Internationale de Philosophie* 27: 224–54.
Russell, B. 1905. "On Denoting." *Mind New Series* 14: 479-93.
Russell, B. 1918. "The Philosophy of Logical Atomism." *The Monist* 29: 190—222.
Walton, K. 1990. *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge: Harvard University Press.

Croatian Journal of Philosophy
Vol. XX, No. 60, 2020

Semantic Deference and Groundedness

ANTONIN THUNS
Centre de recherche en linguistique LaDisco, Université Libre de Bruxelles, Belgium

Semantic deference allows for the meaning of a word w *a speaker uses to be determined by the way other speakers would understand or use* w. *That semantic deference has some role to play in semantic content attributions is intuitive enough. Nevertheless, the exact conditions under which semantic deference takes place are still open for discussion. A key issue that the article critically examines is Recanati's requirement that deferential uses be* grounded, *that is, that deferential uses be linked to* non-deferential *uses (Recanati 1997; 2000). After distinguishing between semantic and epistemic deference, I submit that the only way to maintain the Groundedness Thesis for truly semantic deference is to allow deference to idealized future linguistic collectives. I conclude that this is too high a price to pay for Groundedness and I suggest that it should be rejected as a semantic thesis.*

Keywords: Semantic deference, word meaning, grounding.

1. *Semantic deference and the "Groundedness Thesis"*

Speakers defer semantically when the meaning of words and expressions they use is determined by other speakers' understanding of these words/expressions rather than by their own. Typically, when speakers do not master the semantic content associated with a word *w* they want to use, they defer to an external norm for the fixation of the meaning of *w*. The deferential mechanism is meant to account for the important fact that speakers can express meaningful thoughts involving lexical items they have an incomplete or imperfect understanding of. They do this by relying on the lexical competence and knowledge of other speakers.

Semantic deference seems to be supported by fairly robust intuitions about truth-conditions. It makes content ascriptions possible independently of speakers' lexical mastery and also represents the implicit commitments which go along with the use of a public language. It is indeed very natural to attribute its standard meaning to a word *w*

a speaker S utters, even if S's understanding of *w* is such that S would not be judged competent with *w* according to communal standards. If you overhear someone saying that 'The White House is threatened with subpoena', the most natural assumption is that the truth-conditions of this random speaker's assertion do not so much depend on her private understanding of the term 'subpoena' as on the way legal experts would understand it. By default, you assume that the speaker defers semantically to legal experts for the fixation of the meaning of 'subpoena'.[1] The fact that the speaker could misunderstand this technical expression does not make the truth-conditions of her utterance different from the truth-conditions the same sentence would have if it was uttered by you, or by a legal expert, in the same circumstances. Exactly like you or the legal expert, the casual speaker is entitled to *mean* subpoena by 'subpoena', in the double sense that she can not only *aim at* the relevant technical notion but effectively *summon* it in her discourse. This remarkable feat is supposedly achieved through semantic deference.

As Marconi observes, semantic deference is "real enough", and simply corresponds to the fact that "we do not regard ourselves as semantic dictators, like Lewis Carroll's Humpty Dumpty" (Marconi 1997: 90). As members of a linguistic community, speakers implicitly know that a lot more is going on about word meanings than what their limited perspective gives them access to. Language users are aware of the fact that meanings are not entirely up to them. Empirically, this is evidenced by speakers' willingness to have their language use corrected. It is also attested by a tendency to hold speakers accountable for the meaning of words they use, without regard for what they have in mind when using them. If I say there are 'courgettes' in the fridge, I make a false assertion if there happen to be only cucumbers there, even though it is precisely *these* cucumbers I had in mind (I bought them and put them there, I just tend to get confused about the labels). I could argue all I want, I asserted that there were courgettes in the fridge, not cucumbers.

So, however speculative the notion might appear at first sight, strong intuitions support the existence of a mechanism of semantic deference. These can be gathered both from the internal perspective of speakers and from the external perspective of hearers and other "content ascribers". Speakers are disposed to be corrected about word meanings and, even more interestingly, are disposed to recognize that, by the mere fact of using certain words, they express semantic contents they might not have internalized. Accordingly, hearers or other external observers standardly ascribe semantic content to words a speaker utters without regard for the possible idiosyncrasies in the speaker's lexical-semantic representations.

Clearly, the theory of semantic deference meshes with the "anti-individualist revolution" (Bochner 2014) of the 1970s, whose deep lesson

[1] From here on, the formula "S defers to *x* for *w*" will be used as a shorthand for "S defers to *x* for the fixation of the meaning of the word/expression *w*".

was that individual mental states do not always determine linguistic reference. Though Putnam does not use the term 'deference' in "The meaning of 'meaning'" (1975), it is quite obvious that his famous point about the "division of linguistic labor" supports the idea of semantic deference. Like Putnam, you might be unable to tell the difference between elms and beeches, but the mere existence of the two labels 'elm' and 'beech' is evidence enough for you to suppose that two distinct tree concepts are involved, which more expert speakers master and are able to apply correctly. The same is true for 'gold', 'molybdenum' or even, as Putnam argued, for a perfectly common word like 'water', whose real extension-determining "meaning" H_2O is best defined by chemists and thus reaches beyond the superficial understanding that ordinary speakers might have of the word independently of scientific knowledge.[2] Burge (1979) made a similar theoretical point about social kinds such as 'sofa' and 'contract': these terms' meanings depend on the socio-linguistic environment, not on the private conceptions of individual speakers. Burge's "anti-individualism" also provides support for the idea of semantic deference. Generally, semantic deference fits in nicely within an externalist framework. Even better, it tends to strengthen such a framework by showing that ordinary speakers themselves are practical externalists. They seem indeed to go by the assumption that a great deal of semantic knowledge can be safely "outsourced". In the division of linguistic labor we trust.

As appears from these considerations, semantic deference is tied analytically to the notions of knowledge and expertise. If speakers defer in the first place, it is because other speakers have a better grasp of the meaning of some terms and are experts at using them. Expert speakers provide standards of lexical competence. They are needed for the mechanism of semantic deference to work. This requirement is made explicit in Recanati's formalization of the deferential mechanism. Recanati (1997) introduced a "deferential operator" which allows construing "deferential concepts"[3] in a way analogous to indexicals. The deferential operator "$R_x(\,)$" is a function that applies to a "public symbol *o*" (the word which requires deference to take place) whose content is that of the concept possessed by a "cognitive agent *x*" (the expert) and yields a deferential concept "$R_x(o)$" (the one available for the deferrer to use). In Kaplanian terms, the content of the deferential concept thus obtained is the same as the content of the non-deferential concept (the expert's concept), but its character is different. The character of the deferential concept is metalinguistic. Informally, it is something like "whatever *x* means by *o*". So, when the random speaker uses the word

[2] This claim needs to be qualified. For Putnam, it is ultimately water itself, the actual stuff, and not chemists, that determines the meaning of 'water'. This important point will be discussed below.

[3] The use of the term 'concept' by Recanati might be somewhat puzzling. As I take it, 'concept' here refers to the general meaning of a term as it is understood by a particular speaker, that is, the *character* this term has for the speaker.

'subpoena', she implicitly points toward experts in the linguistic community and appeals to their semantic competence. For a deferential speaker, then, the character of the word 'subpoena' should be something like "whatever 'subpoena' means for the relevant expert". But its content (i.e., its Kaplanian "intension") is simply the content it has for the expert.

Experts are supposed to be part of the deferential speakers' environment. As such, they are part of the linguistic context of deferential utterances. The *x* variable in the formula can be taken to be a contribution of the context to the utterance content – hence the analogy with indexicals. Deferential cases are cases of "social indexicality" (Recanati 2001). If there is nobody in the linguistic community to defer to, or if the "experts" themselves do not really understand what a term means, the deferential concept fails to acquire a determinate semantic content, very much like a demonstrative fails to refer in the absence of a demonstrated object. This means that a deferential concept can have a definite character but no definite content or intension, if it appears that in fact no one is in a position to use the term non-deferentially. Woodfield (2000: 442) provides the example of members of a religious sect who believe that "Jesus Christ and John Lennon are *alpha-enlightened* beings", in a situation where no one (not even the guru) is able to really understand the phrase 'alpha-enlightened'. In this case, the deferential operator applies to 'alpha-enlightened', but the deferential mechanism fails to yield any determinate (truth-evaluable) content, because no genuine expert is ultimately available to provide an interpretation for 'alpha-enlightened'. This means that, in the actual world, there is no context that would endow the term with evaluable content. As a result, the assertion of the sect's belief is semantically ill-formed and its truth-conditions cannot be calculated. By contrast, a student who acquires from her teacher a belief that "Cicero's prose is full of 'synecdoches'" (Recanati 1997: 86) comes to entertain a functioning deferential concept of 'synecdoche', whose content can be traced back to a real expert.

Such a constraint on what counts as successful semantic deference constitutes what Recanati calls the "Groundedness Thesis":

> *(Groundedness Thesis)*
>
> A deferential use is grounded only if someone at the other end of the deferential chain uses the expression in a non-deferential manner. (Recanati 2000: 452)

Deferential uses, Recanati claims, are parasitic on non-deferential uses. This means that deference must stop at some point. If everybody uses a word deferentially, then there is no real content associated with the word. It remains empty, meaningless. The Groundedness requisite can then be used to set apart functioning deferential uses from cases in which everybody defers to an authoritative interpretation which is in fact nowhere to be found. The link between semantic deference and se-

mantic expertise, together with the Groundedness Thesis (henceforth, GT) it leads to, constitutes the most natural account of the phenomenon under consideration, which I will refer to as the "standard account". While I think the idea motivating the account is initially sound – because standards of lexical expertise do exist and can be identified – it raises a number of problems. These problems, I argue, call for severe revisions of the notion of Groundedness that end up making it unappealing.

2. *Epistemic deference, semantic deference and collective grounding*

The main problem of the standard account, Woodfield (2000) argues, is that it equates semantic deference with *meaning borrowing* from particular individuals. The application of the deferential operator to an expression *σ* seems indeed to allow the deferential speaker to "borrow" the meaning *σ* has for a particular expert (the one who happens to be related to the deferrer by a deferential chain). But what if the expert gets it wrong? And what if experts disagree among themselves? As Woodfield shows, this can lead to highly counter-intuitive misunderstandings between speakers of the same language, once several experts are involved in a deferential process. In the synecdoche example adapted from Recanati (1997), it is enough to imagine the following dialogue between Alf, the boy who picked up the word 'synecdoche' from his schoolteacher (the initial expert), and a linguist L (the new expert):

> (i) Alf says: 'Cicero's prose is full of synecdoches.'
> (ii) L replies: 'No it is not. It's true that his prose is full of figures of speech. But very few of them are synecdoches.'
> (iii) Alf replies: 'I accept what you say. Cicero's prose is not full of synecdoches.'
>
> (Woodfield 2000: 448)

Let's assume, with Woodfield, that the teacher to whom Alf defers himself has an imperfect understanding of the concept of synecdoche: he uses the term to refer to metonymies (of which synecdoches normally form a sub-kind). If we take the deferential operator to allow Alf to borrow the schoolteacher's concept, then, when he says that 'Cicero's prose is full of synecdoches', Alf is in fact asserting that Cicero's prose is full of *metonymies*, since the operator forces us to treat the teacher's understanding of the term as providing the semantic content of the deferential concept. But, in that case, the expert linguist L, who masters the meaning of 'synecdoche' and challenges the student's assertion, would be in fact missing the point entirely. L is referring to synecdoches, whereas Alf, unbeknownst to himself, is referring to another notion that he borrowed from his teacher under the misleading label 'synecdoche'. Since they are not co-referring, Alf and L are just talking at cross-purposes. Things become even more troublesome if we consider

the possibility that Alf's deferential pointer changed targets between (i) and (iii). If, after accepting L's judgement, Alf immediately stops deferring to the schoolteacher and starts deferring to L instead, then (i) and (iii) no longer stand in logical contradiction to one another, discursive appearances notwithstanding. At the level of propositional contents, the dialogue would indeed run as follows:

(i) Cicero's prose is full of *metonymies*.
(ii) No it is not. It's true that his prose is full of figures of speech. But very few of them are *synecdoches*.
(iii) I accept what you say. Cicero's prose is not full of *synecdoches*.

Obviously, this misrepresents the communicative exchange, as it runs against our strong intuition that Alf and L do mean the same thing all along when they use the term 'synecdoche' (assuming they are engaged in a cooperative linguistic exchange), even if they do not understand it the same way and even if Alf's deferential attitude changes in the course of the exchange. Furthermore, it misrepresents the fact that Alf has apparently changed his mind about one and the same topic between (i) and (iii), namely, about synecdoches in Cicero's prose.

Fortunately, this problem can be overcome by drawing, with De Brabanter et al. (2005), a distinction between "semantic deference" and "epistemic deference".[4] *Epistemic deference* is about directly exploiting others' cognitive resources. We defer epistemically to other people when we assume that they have a better understanding of the reasons and evidence for making certain claims we make. An agent A_1 who defers epistemically to another agent A_2 uses A_2's words and behavior as reasons to make certain claims and draw certain conclusions, without necessarily understanding the criteria and reasonings motivating A_2's claims and behavior. Epistemic deference is not a specifically linguistic phenomenon. As Rauti (2012: 326) points out, you may defer epistemically to a silent hunter in order to know which way to turn in a hunt for deer. Whether the hunter comments or not on the signs she spotted is irrelevant to your ability to defer epistemically to her. You do it as long as you make your own a claim or a belief based on someone else's knowledge or expertise, without thereby necessarily acquiring this agent's knowledge or expertise.

Semantic deference, by contrast, is a specifically linguistic phenomenon. It concerns the fixation of word meanings themselves, i.e., the fixation of their stable semantic contribution. A speaker S_1 may defer epistemically to another speaker S_2 for the reasons to use a word in a

[4] In their taxonomic study, De Brabanter et al. (2005) clarify the distinction not only between semantic and epistemic deference, but also, within semantic deference, between default and deliberate deference. They also show that the problem of imperfect mastery must be considered separately, as it is not a necessary feature of deferential uses (one can deliberately defer to someone else's use/understanding of a word even in the absence of imperfect mastery, say, for playful or ironic purposes). I refer the reader to their work for further clarifications.

certain way. But it does not entail that S_1 defers *semantically* to S_2 for the very meaning of that word. As De Brabanter et al. (2005) show with their adaptation of Burge's arthritis example, a patient can defer epistemically to her doctor concerning the applicability of the medical term 'arthritis' to her condition. But she does not thereby defer semantically to this particular doctor for the fixation of the meaning of the word 'arthritis' itself. Even more, she cannot reasonably be expected to defer semantically to him.

Let's imagine that she is correctly diagnosed with arthritis by her physician, but that it appears later that he actually has a poor understanding of the ailment and, though able to diagnose it correctly (because he is good at identifying its symptoms), is not a reliable theoretical guide to arthritis (he holds many false beliefs about the ailment). Would the patient be willing to admit she was deferring all along and in every regard to her doctor for the *meaning* of 'arthritis', say, each time she was talking to people about her condition? We have reasons to doubt it. Even if her doctor was her only source of information about arthritis, it does not seem to imply she meant everything her doctor means by 'arthritis' each time she used the term. What she meant to refer to was the ailment that is supposed to be described by modern medicine under the label 'arthritis'. Consequently, it is more reasonable to assume that the patient, *by default*, defers semantically to the community of medical experts, and only defers epistemically to her doctor for practical purposes.

Coming back to the synecdoche example, it is easy to see that Alf and the expert linguist both defer semantically to the community of experts in rhetoric. They are definitely talking about the same thing: the semantic contribution of the word 'synecdoche' remains stable over the course of the dialogue. The difference between Alf's use and the linguist's use of 'synecdoche' is purely epistemic. Alf starts off deferring epistemically, first to his teacher, then to the linguist. He takes their claims as reasons to form particular beliefs involving the 'synecdoche' concept. The linguist, here, does not defer epistemically to anyone. But it does not mean that she would never revise her understanding of 'synecdoche', e.g., in the light of new developments in the field of rhetoric. She is still *disposed to defer*. So, if one accepts that default semantic deference consists precisely in this default disposition to defer, both Alf and the linguist defer semantically, by default, to the community of experts in rhetoric, which provides the ultimate norm regarding the correct use of the word 'synecdoche'.[5]

[5] To be fair, this move is already implicit in Recanati (1997: 85), since he also writes that "there is a public interpretation of 'synecdoche', on which experts in rhetorics converge". What we have seen is that this conflicts with his other claim that "the symbol's content, in this context, is the content which the symbol 'synecdoche' has for the teacher" (92) because the teacher, as an isolated expert, could very well be mistaken about the norm.

Of course, epistemic deference may also have linguistic implications. If A_1 uses A_2's *words* as a ground for epistemic deference, it may have an impact on the way A_1 conceives of the *meaning* of some of A_2's words. If someone points out a tree to you and tells you 'That's an elm', you may or not decide to defer epistemically to that person. If you do, you make that person's claim your own and you can use it to tell someone else, pointing to the same tree, that 'That's an elm', or even to tell anyone else later that 'There was an elm there'. Obviously, this process might help you get a better grasp of the meaning of the word 'elm'. Likewise, Burge's Bert, the man who believes that arthritis can affect soft tissues as well as the joints (Bert believes that arthritis has lodged in his thigh), comes to a better understanding of the meaning of 'arthritis' once he is "apprised of the fact that arthritis is an inflammation of joints and cannot occur in the thigh" (Burge 1979: 198). Arguably, Bert may have come to this improved understanding of the meaning of 'arthritis' by deferring epistemically to other people's judgment involving the word 'arthritis', i.e., by making their claims his own without necessarily understanding the rationale behind them. So, in both cases ('elm' and 'arthritis'), it appears that a process of epistemic deference can have consequences on the deferential agent's lexical representations. But, as the discussion of the aforementioned examples is meant to show, it does not follow that epistemic deference has any role to play in the *fixation* of word meanings themselves. From the point of view of semantic content attributions based on deferential uses, local processes of epistemic deference are an idle wheel. Only *semantic-deferential* uses are the legitimate objects of GT.

As Woodfield notes, it is true that real-life processes of deference may stop at particular experts, but it is not true that they have to. An expert A at the end of a deferential chain could always herself defer for some reason to other experts B and C, while B and C defer to A for different reasons (Woodfield takes the example of scientists working together on a disease). Even if A stops deferring to B and C, another expert D can always show up with new evidence and reconfigure the deferential relations. It is then possible to imagine never-ending "*circles* of mutually deferring agents" (Woodfield 2000: 435). The result is that no one is in a position to use a given technical term in a completely non-deferential way, which contradicts GT.

This should not be a problem. Once semantic deference to the "first expert at hand" (De Brabanter et al. 2005: 14) is abandoned and the distinction between epistemic and semantic deference is introduced, the idea of mutually deferential collectives starts to make a lot of sense. Real-life experts only provide partial and fallible ways to approach the meaning of a word. Laypeople defer to them epistemically, not semantically. Likewise, experts may defer epistemically to one another in various ways, but the existence of ideal *individual* experts, worthy of full semantic deference, is highly implausible and maybe simply not required for *semantic* purposes.

The conclusion of this section is that, if semantic-deferential uses have to be grounded, the only plausible "cognitive agents" for the deferential operator to index are *linguistic collectives* made up of variously competent and mutually deferential speakers. This, it seems, follows naturally from the distinction between epistemic and semantic deference. Now, this requires making the notion of a linguistic collective precise and fit to play the role of the "cognitive agent" demanded by the deferential operator. What, exactly, is the "collective" target of the deferential operator? Two options are open. The operator could be targeting only the best experts of the field. Even if none of them individually possesses all the available knowledge, their combined knowledge might deliver the best possible definition of a term. Alternatively, the operator could be targeting something more abstract: a common body of knowledge, a set of norms or even a public language taken as a whole.[6] Perhaps the second option fits better with our intuition that common knowledge is more than the mere sum of individuals' knowledge and is exempt from idiosyncrasies. Moreover, parts of it could even not be known by anybody but nevertheless available, stored in books or, as is increasingly the case, in computers.

To be sure, the notion of an abstract linguistic standard is hard to pin down and might raise ontological questions.[7] Committing to such a notion, however, seems to be a first price to pay if one wants to maintain GT for truly *semantic* deference (as opposed to epistemic deference). Yet, as I am about to argue, even clarifying this notion would not get proponents of GT out of trouble. Indeed, in order to maintain GT, one must make sure that collectives *in fact* play the role of higher-order experts stabilizing meanings. So far, the discussion of 'arthritis' and 'synecdoche' only shows that we are inclined to think that they *should* do so. Yet it does not mean that they do. As the next section purports to show, the strong ties between semantic deference and externalism indicate that they do not.

[6] All these suggestions are found in De Brabanter et al. (2005): "[The] truth value [of an utterance of 'I have arthritis'] is determined by appealing to the experts, and to the linguistic community more generally, regarding the question of what counts as 'arthritis'" (4); "In English, the meaning of "arthritis" is established in connection with the common body of medical knowledge" (5); "[The arthritis patient] is deferring by default to the norms of the linguistic community" (9); "we will be using the variable x either for users of a language or for the language itself" (8).

[7] As O'Madagain (2014) points out, in order for the conventional semantic value of a term to be determined by a group, groups must be susceptible of being attributed intentional states and concepts. However, the model of group semantics he offers, inspired by that of Lewis (1969) for conventions, only works for attributions of intentional states to groups that are the result of an explicit procedure (as is the definition of the term 'meter' by the International Bureau of Weights and Measurements). This account thus falls short of explaining how distributed knowledge could ground the semantics of terms whose meaning is not fully determined by explicit agreement.

3. *Deferring beyond current understanding: virtual grounding*

The deep problem with the "standard account" of semantic deference, even in the amended version where abstract collectives replace individuals, is that it rests on descriptivist assumptions. Saying that speakers defer to experts' *understanding* of a term is tantamount to saying that speakers defer to the experts' *definition* of that term. Once we accomplish the move to abstract linguistic standards, the underlying assumption is that the community provides *the most complete definition*, while no individual expert speaker possesses all the relevant information. Yet, at least in the case of natural kind terms, descriptivism is usually held to be untenable, because questions about a term's meaning can hardly be separated from questions about our epistemic access to the term's denotata. Given our epistemic fallibility, the semantic import of natural-kind term is taken to reach beyond available descriptions of the denotata and, even, to be independent of them. What, indeed, if the experts collectively got some facts wrong concerning the referents of natural kind terms, so that no appeal to collective wisdom can compensate for individuals' shortcomings? Can collective misconceptions ground meanings in the desired sense? Even if collective conceptions are not entirely mistaken, they might simply fall short of providing the determinacy of meaning that we take natural kind terms to have. And how are we to deal with radical changes of conception that occur in the course of a term's history (for example, the fact that fire is now thought of as the effect of a chemical reaction and no longer as a fundamental element)? These sorts of questions lead, from the rather uncontroversial observations that speakers defer and that there is division of linguistic labor, to more radical claims about meaning externalism.

As Putnam famously argued, the fact that our conceptions of natural kinds change does not entail that the *meaning* of natural kind terms changes accordingly. His point was precisely that the meaning of natural kind terms depends on the external world itself, independently of anyone's grasp of the corresponding referents, hence independently of available *descriptions* of those referents. As Liu (2002) highlights, Putnam's thought experiments about natural kind terms and Twin-Earth aim above all to establish the doctrine of *physical externalism*, according to which the meanings of natural kind terms are determined by objective natural boundaries, independently of the state of our collective knowledge. "'[M]eanings' just ain't in the head!" (Putnam 1975: 227) implies that meanings are external to any head, society's heads included. Even so-called experts do not determine meanings. At most, they only provide the best approximations of how meanings should be construed. Therefore, experts, even collectively, do not ultimately ground our use of natural kind terms like 'elm', 'beech', 'water' or 'fire'. Only nature does. In the face of this, it is fallacious to jump from Putnam's "division of linguistic labor" to the standard account of semantic

deference in which experts ground meanings. It is, for example, fallacious to attribute to Putnam, as I did above, the view that chemists determine the meaning of 'water'. Chemistry is at best the most refined method to approach the nature of that which we call 'water', but the meaning of the word, according to Putnam, depends on the kind of substance water actually happens to be, without regard for our conceptions. Twin-Earth-style thought experiments thus imply a rejection of meaning descriptivism for natural-kind terms.

If these externalistic intuitions about meaning are correct, then the conditions under which GT can be maintained must be amended further. Speakers must be taken to defer semantically *beyond current standards of expertise and understanding*. What, in this case, are they deferring to? The intuition behind physical externalism seems to be that speakers must be taken to defer to "the world", to nature itself and the substances it contains. Though it accounts for truth-conditional intuitions, this construal of semantic deference is a considerable departure from the original notion of deference and one might object that it does not deserve the label "deference" at all (De Brabanter and Leclercq, ms). However, I think that it is possible to help ourselves to Peirce's conception of knowledge (Peirce 1878) in order to recast this "semantic deference to the world" in a way that preserves the sociolinguistic nature of deference (speakers defer to other speakers) and the descriptivism underlying the reconstruction of the phenomenon. This can be done by allowing speakers to defer to *idealized future collectives of experts*. According to this interpretation, the meaning of the words 'fire' or 'water' is determined not by what current collectives of experts happen to agree on, but by what future collectives will discover about the "substances" in question.[8] So, instead of a rather metaphorical deference to the world, one can maintain the basic notion of deference to other speakers, except that these now are hypothetical future speakers that we take, by stipulation, to be "ideal" experts, in possession of an understanding/definition so perfect that it is as good as the world itself to determine the extension-determining meaning required by our truth-conditional intuitions (those motivating semantic deference in the first place). This kind of *counterfactual* or *virtual deference* (deference to virtual ideal speakers) is a way of neutralizing the undesirable consequences of the descriptivism underlying talk of semantics deference whilst maintaining the structure of the standard account as well as GT. Obviously, since "cognitive agent *x*" is now located in the future, the amended account leads to the introduction of a notable *teleological* dimension in Groundedness. Paraphrasing Peirce's formula, one could summarize the account by claiming that the meaning of that which a word encodes does depend on the real fact that investigation

[8] The proposal is a form of "Temporal Externalism" in the sense of Jackman (2005), except that it is augmented by the necessary idealization of future collectives (see below).

is destined to lead, at last, if continued long enough, to stabilizing that meaning.[9]

However, the Peircean solution only seems to apply to alleged natural-kind terms: a notion that is notoriously difficult to clarify. One might indeed object that words like 'water', 'elm', 'fire' and 'arthritis' all designate "natural kinds" only in a very broad and unenlightening sense of the term. As to physical externalism, it faces the "*Qua*-problem" (Devitt and Sterelny 1999: 79): How can we know that speakers, upon naming samples of a given entity that we *retrospectively* interpret with the categories of contemporary science ('Water has a chemical formula'; 'Elms form a genus'; 'Fire is a chemical reaction'; etc.), actually converge on the implicit norm of the hidden nature affording the natural-kind interpretation? Whilst acknowledging the difficulty, I will not engage in any epistemological discussion about the grounding of natural kind terms here. As I now will argue, the point on "deference beyond current standards" is much more general and does not crucially depend on the outcome of such a discussion. It can even be made by considering terms with a far more dubious epistemological status than alleged natural kinds.

Take *polemical terms* such as 'justice', 'science', 'philosophy', 'democracy'. They pose a specific problem for the standard deferential account, because their meaning is still so heavily debated in the community of experts, without any consensus emerging, that they suffer, as they stand, from a lack of determinacy which threatens the propositionality of the sentences that contain them, hence the possibility for semantic deference to be grounded. Most philosophical and moral terms are essentially polemical. Concepts debated within the social sciences and linguistics are also often polemical. Just consider, to take but one example, the issue of the distinction between what counts as 'semantic' and what counts as 'pragmatic'. The range of conflicting views available is such that it could be argued, by a purely external observer endorsing the Groundedness principle, that these alleged terms of art are in fact deprived of meaning. I take it that a similar point could be made about the need to distinguish between 'synecdoche' and 'metonymy'.

Yet, experts disagree precisely about what they take to be one and the same conceptual problem, which they represent by using the same expression. They consider that there should be only one version of the 'semantic/pragmatic' distinction,[10] or one clear-cut definition of 'syn-

[9] "'Truth crushed to earth shall rise again', and the opinion which would finally result from investigation does not depend on how anybody may actually think. But the reality of that which is real does depend on the real fact that investigation is destined to lead, at last, if continued long enough, to a belief in it." (Peirce 1878: 274)

[10] This claim is compatible with the facts that experts currently recognize that there are many theoretically sound ways of making the distinction. I think that this is due to our epistemic limitations. At the end of the day, the presumption behind the use of dichotomic scientific labels is that there should be a clean way to "draw the line". If there isn't, then the dichotomy should be abandoned or reformulated.

ecdoche', because these are supposed to capture something about the investigated objects themselves. In the same vein, many philosophers assume that there is something substantial to our intuitions about 'morality' and 'justice', or about the autonomy of the 'political', and engage in passionate discussions about these topics, whereas the very existence of valid objects of inquiry has not been settled yet (and, in these cases at least, cannot be settled simply by pointing at natural samples of the referents, as physical externalism would have it). We can have the intuition of the unicity of an object/topic and coin a term for it long before we know whether the intuition is justified in the first place. In such cases, a deferential attitude appears to be almost constitutive of, and built into, the use of the term.[11] This means that there should be some kind of external linguistic standard to which speakers are taken to defer by default. Given the currently polemical nature of these terms, the linguistic standard that speakers are gesturing toward only comes as a promise, as the possibility of a future agreement.

Let's recapitulate. Whether we consider natural kind terms or polemical terms, the point is basically the same: their semantic content is not plausibly stabilized – hence not grounded – by current collectives of experts. Yet we have the intuition they should be grounded for the sake of semantic content attributions. Therefore, GT has to be amended in the proposed way: by allowing default semantic deference to idealized future collective states. Deference to idealized future states of the community reflects the situation in which, on the one hand, there is some implicit agreement that there should be a stable meaning behind a given term and, on the other hand, the community of experts is unable to ground it, either because the extension-determining meaning of the term depends on the hidden nature of its referent, which only a hypothetical perfect state of knowledge could fully reveal (natural-kind terms), or because its meaning depends on a hypothetical perfect coordination between future experts (polemical terms).

This argument, I believe, works for all terms whose meanings have a teleological component. In our proposed Peircean construal, this concerns natural-kind terms as well as polemical terms. *Teleological terms* refer to entities our understanding of which is always perfectible and which are used in a process of ever-growing knowledge (natural-kind terms) or ever-finer rational argumentation (polemical terms). Follow-

[11] Kaufmann (2006) explores the idea that "socio-political concepts" (expressed by terms like 'nation', 'God', 'public opinion') are constitutively deferential because they do not exist independently of the communicative chains through which users get acquainted with them. The way they are given is purely communicational. Ordinary agents gain access to them only by deferring to higher authorities (Church, political elites, social scientists) who also happen to use the terms deferentially. Experts defer to higher authorities or principles, which are only evidenced through language (writings, oral traditions, testimonies), and, crucially, never through the provision of an extra-linguistic referent. The deferential chain never stops: socio-political concepts only have a reference "on credit". As such, they blatantly violate the Groundedness principle in its original formulation.

ing Jackman (2005), one may *characterize* (in the technical sense of *character*) the meaning of teleological terms as being sensitive to "epistemic" considerations. This means that changes in our conceptions of the meaning of these terms are brought about by epistemic factors: new discoveries or a better organization of our knowledge. Jackman contrasts "epistemic" with "pragmatic" explanations of linguistic change:

> "Pragmatic" theories explain change in terms of nonrational factors such as taboo, metaphor, semantic drift and the like. For instance, 'zipper' changes from being a brand name to a generic name for such mechanical fasteners because the brand is so successful that users of the language gradually forget that the items of that kind are ever called anything else. We have a sociological/psychological explanation of the change, but no justification in terms of the truth of the beliefs involved. By contrast, "epistemic" theories explain changes in usage in terms of factors such as the need to keep our beliefs consistent both with new experience and with each other. We stopped, say, applying the term 'fish' to whales because we discovered that whales were in many important respects closer to those creatures we called 'mammals' than to other creatures we called 'fish'. (Jackman 2005: 370)

In Jackman's view, we take epistemically-driven changes to already affect what we mean with teleological terms, whether we can foresee these changes or not (and whether speakers explicitly acknowledge being "temporal externalists" or not). In the Peircean version of temporal externalism I claim proponents of GT are committed to, laypeople, individual experts and currently existing linguistic collectives use teleological terms deferentially. They all defer beyond the limited horizon of the community's current epistemic state relative to the concepts at stake. They defer to a more advanced state of the debate, to an ideal state of the epistemic community in which a critical mass of interpretations would converge. What matters for semantic content ascriptions is not the actual criteria speakers use, but the fact that they are collectively engaged in using a term with the presumption that it corresponds to a valid concept. As a result, the ultimate semantic import of the term is more virtual than actual and calls for a notion of *virtual semantic deference*, i.e., deference to a potential or future normative agreement on the meaning of the term. Committing to such a notion, I suggest, is the price to pay if GT is to be maintained regarding deferential uses of teleological terms.[12]

[12] Schroeter and Schroeter (2014) propose a "connectedness model" purporting to solve the question of the semantic content of "normative concepts" along very similar lines. Their model abandons the requirement of sameness of criteria and replaces it with a "tradition-based determination theory" of semantic values: the "entire representational tradition (*i.e.*, the entire set of token thought elements bound together by relations of apparent *de jure* sameness)" (2014: 12) is taken as the default semantic value for the word/concept. Historically extended representational traditions encompass conflicting views and could therefore be taken as the default value of what I called "polemical concepts". If representational traditions are extended in the past, it seems that they also point toward the future: future epistemic states of the community and future agreements between experts might solve the current lack of referential anchoring of some traditions.

4. *Too tough a bullet to bite?*

I have argued that GT commits the theorist to far more than the existence of expert speakers, as long as the thesis is to apply to semantic deference and not merely to epistemic deference. Semantic deference concerns the ascription of truth-conditions to utterances. Given the implicit externalism of our linguistic practice concerning certain terms, truth-conditions are world-involving and cannot be fully determined by mental states alone, even by the mental states of the most knowledgeable speakers. Abstracting away from individual speakers towards the communal level fits better with what speakers take themselves to be doing, as the 'arthritis' example illustrates. Yet, upon closer scrutiny, deference to more abstract linguistic standards is not enough to vindicate our truth-conditional intuitions. As Jackman (2005) points out, changes in usage guided by epistemic factors (for example, the discovery that whales are not fishes) are normally not seen as changes of meaning, so that we also need to treat future "conceptual developments" as already affecting the meaning of terms we use. The fact that whales have been reclassified as mammals at some point does not seem to entail that the meaning of 'whale' has changed. We have simply acquired a better grasp of the properties that delineate the extension of the term, that is, a better grasp of the extension-determining meaning of the term.[13] So, the intuition is that extension-determining meanings are already there, fully determined, despite the fact that we only very imperfectly apprehend them. Supposing the phenomenon of semantic deference is to retain its socio-linguistic dimension and be grounded at the same time, I have submitted that there is no other option for the deferentialist than to accept that we defer, implicitly and by default, to idealized future states of collective understanding.

Now, my final claim is that the amended account of Groundedness I have just sketched is probably too committal. For one thing, it obviously rests on a very severe idealization. But, much more problematically, *the account abusively attributes a foundational role to intuitions*. This last problem, I think, is serious enough that it is preferable to reject GT altogether and reframe the issue in less demanding terms.

It is one thing to recognize, as Jackman does (2005: 370), that, in the counterfactual situation where we could see into the future and discover an epistemically superior usage of one of our terms, we would be willing to change our own usage accordingly (thereby showing that we take ourselves to be same-meaning with future speakers). It is quite another to assume, as I have done for the sake of dramatization, that some metaphysical connection between our usage and a hypothetical future perfect understanding *presently* grounds the semantic import of teleological terms. Viewed in this light, virtual grounding is nothing

[13] Of course, other, non-denotational aspects of the "meaning" of the term, such as prototypical representations and stereotypes, may have changed as a consequence of the reclassification. But these are not supposed to affect literal truth-conditions.

more than "temporally loaded"[14] Platonism. Yet I claim that this is precisely what GT inevitably leads to, if *semantic* deference is supposed to lead to non-deferential uses.

Perhaps one way to vindicate the intuitions motivating talk of semantic deference whilst avoiding outlandish metaphysical claims is to reject Groundedness as a metaphysical thesis and recast it as a psycholinguistic hypothesis. Put very briefly, the rationale behind this move would be the following. The external world is essentially inscrutable. Likewise, our relation to the external world is inscrutable**.** As a consequence, the extent to which the use of our terms follows precise epistemic norms (e.g., the norm of the "hidden nature") is not known a priori, essentialist intuitions notwithstanding. Admitting that, it is not inconceivable that a great deal of our terms whose meaning is sensitive to epistemic considerations ("teleological terms") are in fact semantically indeterminate: their boundaries evolve with our dealings with the world without being firmly settled. There is no reason to assume, just because we have realistic semantic intuitions, that nature itself or future speakers are systematically cooperating to fully determine the semantic import of the teleological terms we use. However, this does not prevent us from *taking* these terms to have determinate contents, exactly in the same way that we take ourselves to speak English and not just similar idiolects. It is a psycholinguistic fact that teleological terms appear to us to be endowed with substantive content and that, most of the time, a change of our understanding does not appear to us as a change of meaning. In fact, it is possible that we do not know whether the extension-determining meaning of a given term fully obeys "epistemic" rather than "pragmatic" considerations. Consequently, our commitment to the stable semantic import of teleological terms could be, at least in a significant number of cases, the result of an illusion of the semantic kind.

Take the case of the word 'whale'. Most likely, its extension-determining meaning has not changed over the last few centuries, because the term plausibly had the same extension in 18th century fishers' mouths as it has in ours. But let's imagine a counterfactual situation in which ancient fishers had come across some species of very large fish that looked very much like whales. I do not see any reason not to include this species in the extension of 'whale' in their English, if ancient fishers would so call those large fish. This would not be incompatible with the fact that subsequent speakers, once they have discovered better ways of classifying species, would perhaps say that 'in fact those fishes weren't whales' – precisely the kind of argument the temporal externalist appeals to. To be sure, speakers could opt for the exclusion of large fish from the extension of 'whale', thereby abiding by the scientific use. But it does not follow from this that whale-like fish were not in the extension of 'whale' before the discovery of the fact

[14] Jackman's phrase.

that *most* of the animals called 'whales' are mammals. The extension of the term has simply been changed. Therefore, the informal claim that whale-like fish were incorrectly called 'whales' is a *retrospective semantic illusion*. In this scenario, whale-like fish just *were* in the extension of 'whale'. They simply no longer are. Alternatively, in the same counterfactual situation, speakers might just refuse to follow epistemic considerations and keep using the word 'whale' as they used to (i.e., as a term referring to both whales and whale-like fish), leaving to biologists the coinage of other terms for the purpose of precise classification. In this second scenario, whale-like fish unproblematically were and remained in the extension of 'whale'. In fact, something very similar happened in the actual world for the term 'reptile'. Scientifically, this term is no longer deemed to coincide with a consistent natural kind. In ordinary language, birds do not count as 'reptiles', whereas crocodiles, turtles, snakes and lizards do. In cladistic terms, this does not make much sense, for crocodiles are phylogenetically closer to birds than to the other 'reptiles'. In spite of this discovery, 'reptile' has remained a perfectly serviceable English word and its extension has remained the prescientific one. Theoretically, its use could have been extended to include birds, and one could have said that 'in fact, birds are reptiles too', but that is not how usage has evolved. Instead, there is still an ordinary use for 'reptile' (probably motivated by the superficial properties so-called reptiles share, which provide useful contrast with birds, mammals and amphibians, that is, motivated by "pragmatic" factors) and no well-founded scientific use.

I take it that such examples are easy to multiply, showing that externalist intuitions about the "true scientific extension" of a natural-kind term suffer from what one might call a "scientific-essentialist" bias. My point is not that usage is never sensitive to essentialist considerations. It is rather that, even if ordinary usage does end up following scientific norms, it is only a retrospective illusion that the term "always had" the extension-determining meaning and the extension that we now take it to have. A less committal thesis would be that such illusions and other objectivistic intuitions about word meanings are compatible with the *de facto* semantic indeterminacy of a number of teleological terms and the open-ended nature of the semantic deference attached to them. The intuitions on which talk of deference and groundedness is based may have an important psychological role to play and are probably indispensable to the project of knowing the world and communicating about this knowledge. However, by themselves, these intuitions are not enough to *ground* semantic deference, except if one is disposed to embrace a form of teleological Platonism. Unless one is willing to bite that bullet, I take it that GT should be rejected.

5. *Conclusion*

Once linguistic deference is taken to determine *semantic* content proper, it forces us to go beyond the level of concrete interactions, or even chains of interactions, between speakers and to consider more abstract linguistic standards that serve as guiding principles for our semantic content ascriptions. More precisely, only virtual semantic deference to future idealized linguistic collectives seems to be able to secure the intended semantic import for our teleological terms. Barring the explicit endorsement of a form of Platonism, such an account is too committal and is better recast as a psychological hypothesis.

My focus has been mainly on prototypical examples of semantic deference: deference for terms like 'elm', 'water', 'arthritis' and 'synecdoche', all amenable to scientific enquiry. I also addressed the category of polemical terms, which motivates the introduction of the more inclusive category of teleological terms: terms pointing at some "realities" (not necessarily physical) which might never be fully known but to which speakers nonetheless take themselves to be referring through words. I take the category of teleological terms to encompass all the terms whose meaning is sensitive to epistemic considerations: natural kind terms, scientific terms, polemical terms.

Teleological terms provide prime illustrations of semantic deference, because examples involving our epistemic limitations are the ones that come to mind most easily. A question worth exploring, though, is whether speakers can defer semantically for words whose meaning does not fit nicely in the expertise-based model that underlies the discussion here. Besides teleological words, there is a vast grey area of relatively imprecise words, which are nevertheless easily grasped by speakers because of their familiarity. For most of these perfectly ordinary items, it is not clear that it is possible to identify a community of experts, nor a teleological dimension attached to them. Are there specialized bodies within the English-speaking community who stabilize the lexical meaning of the words 'meeting', 'sadness', 'soft' or 'friend'? Is the linguistic community engaged in the pursuit of a better understanding of the semantic import of these terms? Most likely not. Moreover, it is arguably part of the very functions of ordinary non-technical words to remain semantically underdetermined, in order to provide very abstract and flexible frames for the pragmatics of everyday communication to fill in.

Yet these words intuitively express distinct and graspable notions. Even though they are non-technical, vague and extremely context-sensitive, they are not obscurely indeterminate in the sense esoteric concepts are. So, it seems that someone who does not understand an ordinary word like 'friend' (a small child, typically, or a foreigner), could defer semantically to other speakers for such a term. Would this kind of deference be constrained by GT? What would it mean to say that another speaker uses the word 'friend' non-deferentially? Since

the general concept of 'friend' is semantically underdetermined, it can be used with divergent semantic imports by equally competent speakers. Does it mean that each non-deferential speaker grounds 'friend' in their own fashion, so that the concept is "multiply grounded"? This is an interesting possibility, but it runs counter to our understanding of the implicit motivations of the deferential speaker. Someone who defers semantically for the meaning of 'friend' does not want to wind up expressing any old speaker's idiosyncratic take on the word. The deferential speaker, by default, aims at the standard, collective concept. But what is the collective concept of 'friend' and how is it grounded? It does not seem to be a teleological term, because its use is more sensitive to pragmatic rather than epistemic considerations. Knowing that future speakers will use the term differently would not be a reason for us to consider that our current use of 'friend' is mistaken in any way. And yet, what if future psychology converted words like 'friend', 'love', 'trust' or 'doubt' into full-blown natural-kind terms, so that semantic appropriateness of past uses could be reevaluated with respect to new scientific findings?

These reflections are, of course, purely speculative, but the fact that they make sense hints at the positive side of Section 4's conclusion. If, on the one hand, it is possible that many of our teleological terms are semantically indeterminate, it is also possible, on the other hand, that some apparently non-teleological terms are more determinate than we anticipate, because they could also end up falling under epistemic considerations. In a sense, the whole project of the "special sciences" (sociology, psychology, anthropology…) is to unveil the "hidden" social and psychological functions of a host of ordinary terms, i.e. to uncover their externalistic semantics. The fact that the terminology of the special sciences is "contaminated" by ordinary language and always runs the risk of being semantically indeterminate is just a particularly dramatic case of the general risk of semantic indeterminacy affecting all our terms, natural-kind terms included. Once GT is rejected as a semantic thesis, strict boundaries between kinds of terms are no longer needed. The pervasiveness of semantic deference and its psychological importance can be fully acknowledged without us committing to a problematic metaphysics of meaning.*

References

Bochner, G. 2014. "The anti-individualist revolution in the philosophy of language." *Linguistics and Philosophy* 37 (2): 91–120.

Brown, M. 2010. *How I Killed Pluto and Why It Had It Coming*. Spiegel & Grau.

Burge, T. 1979. "Individualism and the mental." In his *Foundations of Mind, Philosophical Essays, Volume 2* (2007). Oxford: Clarendon Press: 100–150.

De Brabanter, P., Stojanovic, I., Villanueva Fernandez, N. and Nicolas, D. 2005. "Deferential Utterances." Retrieved from https://hal.archives-ouvertes.fr/ijn_00000575v1.

De Brabanter, P. and Leclercq, B. Manuscript. "Semantic externalism(s) and semantic deference."

Devitt, M., and Sterelny, K. 1999. *Language and Reality. An Introduction to the Philosophy of Language*. Cambridge: The MIT Press.

Jackman, H. 2005. "Temporal externalism, deference, and our ordinary linguistic practice." *Pacific Philosophical Quarterly* 86: 365–380.

Kaufmann, L. 2006. "Les voies de la déférence: Sur la nature des concepts sociopolitiques. » *Langage et société* 117: 89–116.

Lewis, D. 1969. *Convention*. Cambridge: Harvard University Press.

Liu, J. 2002. "Physical externalism and social externalism: Are they really compatible?" *Journal of Philosophical Research* 27: 381–404.

Marconi, D. 1997. *Lexical Competence*. Cambridge: The MIT Press.

O'Madagain, C. 2014. "Can groups have concepts? Semantics for collective intentions." *Philosophical Issues* 24 (1): 347–363.

Peirce, C. S. 1878. "How to make our ideas clear." In C. J. W. Kloesel (ed.). *Writings of Charles S. Peirce: A Chronological Edition. Volume 3 (1872-1878)* (1986). Indiana University Press: 257–276.

Putnam, H. 1975. "The meaning of 'meaning.'" In his *Mind, Language and Reality. Philosophical Papers*, *Volume 2* (1975). Cambridge: Cambridge University Press: 215–271.

Rauti, A. 2012. "Multiple groundings and deference." *The Philosophical Quarterly* 62 (247): 317–336.

Recanati, F. 1997. "Can we believe what we do not understand?" *Mind and Language*, 12 (1): 84–100.

Recanati, F. 2000. "Deferential concepts: a response to Woodfield." *Mind and Language* 15 (4): 452–464.

Recanati, F. 2001. "Modes of presentation: perceptual v. deferential." In A. Newen, U. Nortmann and R. Stuhlmann-Laeisz (ed.). *Building on Frege: New Essays on Sense, Content, and Concept*. Stanford: CSLI Publications, Stanford University, 10: 197–208.

Schroeter, L., and Schroeter, F. 2014. "Normative concepts: a connectedness model." *Philosophers' Imprint* 14 (25): 1–26.

Woodfield, A. 2000. "Reference and deference." *Mind and Language* 15 (4): 433–451.

* My research is funded by the Fonds de la Recherche Scientifique (F.R.S.-FNRS) and is part of the research project *Cognitive transparency of semantic contents, pragmatic determination of reference* (PDR T.0184.16, 2016-2020). I would like to thank all members of the project for their very helpful comments on earlier versions of the article.

Croatian Journal of Philosophy
Vol. XX, No. 60, 2020

Representationalism, Double Vision, and Afterimages: A Response to Işık Sarıhan

RENÉ JAGNOW
University of Georgia, Georgia, USA

In his paper "Double Vision, Phosphenes and Afterimages: Non-Endorsed Representations rather than Non-Representational Qualia," Işık Sarıhan addresses the debate between strong representationalists and qualia theorists (Sarıhan 2020). He argues that qualia theorists like Ned Block and Amy Kind who cite double-vision, afterimages, etc., as evidence for the existence of qualia are mistaken about the actual nature of these states. According to Sarıhan, these authors confuse the fact that these states are non-endorsed representational states with the fact that they are at least partly non-representational. I argue that Sarıhan's argument contains gaps that suggest that he misidentifies the mistake that leads these qualia theorists to their conclusion. In my view, these qualia theorists do not confuse the fact that the states in question are non-endorsed states with the fact that they are non-representational, but rather mistake certain representational contents, or certain aspects of these contents, for qualia.

Keywords: Perception, representationalism, qualia, afterimages, double vision.

1. *Introduction*

In his paper "Double Vision, Phosphenes and Afterimages: Non-Endorsed Representations rather than Non-Representational Qualia," Işık Sarıhan addresses the debate between strong representationalists and qualia theorists (Sarıhan 2020).[1] Strong representationalists hold that the phenomenal character of an experience is identical with a cer-

[1] Strong representationalists include, for example, Tye (1995), Dretske (1995), and Lycan (1996). Qualia theorists include, for example, Boghossian and Velleman (1989), Block (1996), and Kind (2008).

tain type of non-conceptual representational content. As Sarıhan puts this: "Representationalism or intentionalism, in its stronger variety, is the theory that all introspectible qualitative aspects of a conscious experience are qualities that the experience non-conceptually represents the world to have" (Sarıhan 2020, 7). Qualia theorists hold that the phenomenal character of a perceptual experience outstrips its representational content and, as a consequence, is at least partly constituted by qualia. Qualia, in the relevant sense, are non-representational and intrinsic properties of experiences. One important set of arguments in favor of qualia is based on the phenomenology of specific kinds of visual states, such as double-vision, afterimages, blurriness, and experiences of phosphenes. According to these arguments, introspection reveals that it is not possible to characterize the phenomenal character of these kinds of states exclusively in terms of their representational contents.

Focusing mainly on Block and Kind, Sarıhan aims to undermine the cogency of these arguments (Block 1995, Kind 2008).[2] In particular, he argues that those authors who cite double-vision, afterimages, etc., as evidence for the existence of qualia are confused about the actual nature of these states. According to Sarıhan, these states are "fully representational," that is, their phenomenal character is exhausted by their representational contents. Yet, the fact that these states have representational contents that are inconsistent with the typical behavior of physical objects has the consequence that subjects do not endorse them at the cognitive level. Those authors who cite these states as evidence for qualia therefore confuse the fact that they are non-endorsed representational states with the fact that they are at least partly non-representational.[3]

In order to support this conclusion, Sarıhan argues for two main claims. The first claim is that those states that are typically cited in support of qualia are fully representational states that automatically fail to be endorsed at the cognitive level. Call this the *non-endorsement claim*. The second claim is that authors who appeal to these kinds of states in order to argue for the existence of qualia confuse the fact that they are automatically non-endorsed representational states with the fact that they are non-representational. Call this the *confusion claim*.

I find the paper very interesting, and I very much enjoyed reading it and thinking about the issues raised in it. I agree with Sarıhan that the phenomenological considerations regarding double vision, afterimages, etc. do not provide conclusive support for the existence of qualia. I also share Sarıhan's strong representationalist leanings and believe that these kinds of visual states are fully representational. But I think that the main arguments for the non-endorsement claim and the con-

[2] Sarıhan also considers the arguments put forward by Boghossian and Velleman (1989).

[3] As Sarıhan puts it: "these authors themselves describe such states as if they were non-endorsed representations, before making the logically illegitimate move that they are non-representational" (Sarıhan 2020: 8).

fusion claim contain gaps. These gaps suggest that Sarıhan misidentifies the mistake that leads qualia theorists to their conclusion. In my view, qualia theorists like Block and Kind do not confuse the fact that the states in question are non-endorsed states with the fact that they are non-representational, but rather mistake certain representational contents, or certain aspects of these contents, for qualia. As I will try to make clear throughout this paper, our disagreement is grounded in a fundamental difference in how we analyze the contents of the states in question.

I proceed as follows. In section 1, I focus on double vision to show that there is a gap in the argument for the non-endorsement claim. Analyzing this gap suggests that those states of double vision that are typically cited in favor of qualia are not accurately characterized as non-endorsed states. In section 2, I focus on afterimage experiences to argue that authors who cite these states in favor of qualia may not be confused in the way Sarıhan suggests. My discussion in this section also suggests that these authors accept a problematic premise, namely that we can experience visible properties only as instantiated either in material objects or in experiences. In section 3, I provide a sketch of a representationalist account of afterimage experiences in order to show why this assumption might be false. I propose that afterimage experiences represent a special kind of object, namely what Martin has called a "pure visible" (Martin 2012).[4] If this account is on the right track, it would support the conclusion that those authors who cite afterimage experiences as evidence for qualia mistake certain representational contents, namely certain pure visibila, for qualia.

Three brief comments. First, in this paper, I focus exclusively on Sarıhan's most important examples, that is, on double vision and afterimages. I will not address other kinds of visual states, such as blurry vision and phosphenes. In my view, these kinds of states require separate representationalist accounts. Second, the goal of section 3 is to sketch a representationalist account of afterimage experiences that, if correct, explains why Block and Kind mistake certain representational contents for qualia. I cannot give a full defense of this account in this paper, however. Third, qualia theorists are often motivated by theoretical rather than phenomenological considerations. However, in this paper I follow Sarıhan and focus exclusively on phenomenological considerations.

2. *The argument in favor of the non-endorsement claim*

The non-endorsement claim holds that experiences involving double vision, afterimages, blurriness, and the like, are automatically non-endorsed mental states that are fully representational. This claim sub-

[4] Phillips has developed a closely related proposal in great detail (Phillips 2013). He also argues that experiences of afterimages represent pure visibilia. But, as I will make clear later on, my proposal differs in some details from his.

divides into two separate claims, namely (i) the claim that these kinds of states are *fully representational*.[5] A fully representational state is one whose phenomenal character is exhausted by its representational content. And (ii) the claim that these kinds of states are *automatically non-endorsed states*.[6] A non-endorsed state is a state whose representational content is not endorsed at the cognitive level, and the non-endorsement is automatic if it is not a result of a conscious inference on the part of the subject.[7]

In order to support claim (i), the claim that these kinds of states are fully representational, Sarıhan considers a number of scenarios involving double vision, afterimages, phosphenes, and floaters. I will focus here on double vision. I will consider afterimages in the next section.[8] Sarıhan describes the following scenario. Suppose that you are standing on the roadside at night waiting for your bus after a party at which you had a bit too much to drink. As you look down the road, a motorbike approaches. However, since your eyes are a bit out of focus, due to the influence of alcohol, you see two headlights moving towards you in unison. Since it is entirely dark so that you do not see anything else in your visual field, you do not notice that this is a case of double vision.[9] Consequently, you automatically endorse the content of this experience at the cognitive level and form the belief that a car is approaching.[10] Note, if it had been lighter outside, you would have noticed that the entire scene had doubled, which would have prevented your cognitive endorsement of its content.

If I understand Sarıhan correctly, we can state the argument for claim (i) as follows. In those cases in which the subject endorses the

[5] The argument for this claim remains somewhat implicit in the text. But I hope that my representation of it is fair.

[6] Sarıhan defines a non-endorsed state as a state that has non-endorsed representational content. He writes: "A mental state has non-endorsed representational content if it has truth-evaluable content but the subject doesn't take the content as true on a higher, cognitive level" (Sarıhan 2020: 11).

[7] As Sarıhan puts this: "[A] non-endorsement of an automatic sort does not require conscious deliberation on the side of the subject regarding the illusoriness of the experience in question" (Sarıhan 2020: 14).

[8] I believe that my conclusions can be extended to phosphenes and floaters.

[9] I would like to point out that this scenario is somewhat problematic. The scenario's purpose is to convince us that you have a non-veridical visual experience that represents two headlights moving towards you. But, since you are intoxicated, it is possible that you misinterpret the content of a veridical visual experience at the cognitive level. I describe later on in this paper how experiences of double vision may be veridical. For now, I would like to point out that it is possible to modify the scenario in a way that avoids the problem. Suppose, for example, that, unbeknownst to you, we anesthetize some of your ocular muscles, so that you see double. We also present you with a display containing one single dot that does not give you any indications that you see double. In this case, it is plausible to say that your visual experience represents one dot as two.

[10] I simplified here. You may first have to endorse a more basic content, such as the content that two headlights move in unison towards you.

content of a state of double vision like in the motorcycle example, we have no reasons to attribute to this state any non-representational aspects. We can explain the subject's cognitive endorsement completely in terms of ordinary representational contents. Specifically, we can say that a state of double vision represents one thing, or one scene, as two. This is a non-veridical visual representation of the scene in front of the subject, but there is no need to appeal to non-representational aspects in order to account for its phenomenal character.[11] If we assume further that this holds for the contents of all states of double vision, it follows that we have no good reasons for saying that they involve qualia.

However, as it stands, the argument does not actually allow us to conclude that *all* states of double vision, including those that are usually cited in support of qualia, are fully representational. Even if states of the kind mentioned in the motorcycle example are fully representational, it is still possible that other states of double vision are not fully representational. In order to arrive at the conclusion that all states of double vision are fully representational, the argument needs an additional premise. Sarıhan needs to show that all states of double vision represent one thing, or one scene, as two. As far as I can see, Sarıhan does not state this premise explicitly and does not provide an argument for it. This is a gap in the argument because one might argue that those states that are usually cited in support of qualia do not represent one thing, or one scene, as two. Let me explain.

In the case of the motorbike example, it is plausible to say that your visual experience represents two headlights. Since there is only one headlight in front of you, your visual experience is non-veridical. In contrast, when you press against your eyeball, as Boghossian and Velleman suggest, you see the entire scene before your eyes double. Yet, in this case, it is not clear that you see two scenes when you see the scene before your eyes double. In fact, Boghossian and Velleman deny this. Describing a case in which you press against your eyeball while looking at a line of text, they write that "you cannot even force the resulting experience into representing the existence of two lines, even if you try" (Boghossian and Velleman 1989: 87). Boghossian and Velleman do not ask you to convince yourself that you see two lines. Their target is not a cognitive state. Rather, they speak about the phenomenal character of the visual state. It does not look to you visually as if there are two lines on the paper. Since this is a plausible description of the phenomenology of afterimage experiences, we would need an argument for the claim that these states represent one thing, or one scene, as two.

I now turn to the argument for claim (ii), that is, for the claim that states of double vision are automatically non-endorsed states. Let me first say a bit more about the notions of cognitive endorsement and non-endorsement. I will focus on visual states. It seems to me that cog-

[11] Incidentally, so Sarıhan, this might also explain why qualia theorists do not cite such scenarios as evidence for qualia.

nitive endorsement is best understood as a dispositional notion, namely as the disposition to form various beliefs regarding the contents of visual states.[12] For example, when your visual state represents a blue cup in front of you, you will be disposed to endorse statements such as "This is a cup," "This cup is blue," and "The cup is right in front of me." Moreover, this disposition is automatic if it does not require a conscious inference on your part. Automatic cognitive endorsement is defeasible. In the case of the Müller-Lyer Illusion, your visual state represents the two lines as being unequal in length. This does not change when you learn that they are equal. But you will no longer endorse the statement that they are unequal. Automatically non-endorsed visual states function in a different way. The representational contents of these states do not create the disposition to endorse statements about them in the first place. The reason for this is that the contents manifestly conflict with the behavior of physical objects.

Sarıhan's argument for the non-endorsement claim now consists of a description of the contents of the relevant experiences of double vision that explains why they do not dispose their subjects to endorse them.[13] He characterizes these contents as follows.[14] Often times, when you see double, the entire scene doubles. Since the physical world does not suddenly double, you know that you are seeing double. Moreover, when you see double, solid, non-transparent physical objects are often superimposed on each other and appear semitransparent. But since solid, non-transparent physical objects cannot occupy the same space and do not suddenly become semitransparent, you know that you are seeing double. In general, the contents of those states of double vision that are usually cited in favor of qualia conflict with our knowledge about the behavior of physical objects and this explains why we do not endorse them at the cognitive level.

In order to bring out my worries about this argument, it will be helpful to consider again how Boghossian and Velleman describe double vision. I already addressed cases in which you see double when you press against your eyeball. If Boghossian and Velleman are right, you see one line double without, however, seeing two lines. It is therefore plausible that the contents of these states are accurate and dispose you to endorse them at the cognitive level. But Boghossian and Velleman also consider more ordinary cases. They write, "Similarly, you can see nearby objects double by focusing on distant objects behind them,

[12] Sarıhan speaks of a belief-inducing function of visual states (Sarıhan 2020: 8). In my view, this is best understood as the disposition to form beliefs about the content.

[13] We can also understand this argument as an inference to the best explanation. Representationalists can give a better explanation than qualia theorists for why we do not endorse these kinds of states.

[14] According to Sarıhan, this should be taken as speculative psychology. Sarıhan leaves open the possibility that empirical research might lead to a different analysis of the contents of experiences of double vision (Sarıhan 2020: 14).

and yet you cannot get yourself to see the number of nearby objects as doubling" (1989: 94). Here, too, it is plausible that the contents of these states are accurate and dispose you to endorse them at the cognitive level. The following example makes this clearer.

Suppose you are walking through a forest. As you look around, you focus on trees at varying distances. Now suppose you first look at one particular tree that is very close to you and then focus on a tree in roughly the same direction that is farther away. When you do this, you see the tree that is closer to you double, but you do not see two trees. The same holds for the trees in its vicinity. Imagine the opposite were true. We would have to accept that your visual state represents different numbers of trees when you refocus your eyes. Since this is highly implausible from a phenomenological point of view, we should characterize the contents of this state in the way suggested by Boghossian and Velleman. Your visual state represents trees at different distances. When you focus on trees that are farther away, you see the ones that are closer to you double without, however, seeing them as doubling in number. Since you have two eyes that are some distance apart from each other, this is as it should be. In other words, you enjoy an accurate experience of the trees in front of you. Based on this visual experience, you will be disposed to endorse a host of statements, such as the statement that there are trees in front of you, that there are so and so many trees in front of you (if the number is relatively small), and so on.

If what I have said in the previous paragraph is correct, we have good reasons for thinking that the states of double vision cited by Boghossian and Velleman function in an entirely ordinary way. They have accurate contents that are usually endorsed at the cognitive level. These visual states are not automatically non-endorsed states. I would therefore conclude that it is not the case that, as Sarıhan claims, "the phenomenology described by Boghossian and Velleman above is better analyzed as a case of non-endorsed representation of doubleness" (Sarıhan 2020: 16). Boghossian and Velleman do not confuse the fact that these states are non-endorsed with the fact that they are non-representational. In order to convince us otherwise, Sarıhan would have to close the gap in the argument for claim (i) and show that all states of double vision represent one thing, or one scene, as two.[15] I will present a similar argument about afterimage experiences in the next section.

In spite of my criticism of the arguments in favor of the non-endorsement claim, I do not think that Boghossian and Velleman have made a conclusive case for qualia in the cited paper. My argument so far suggests that we need to distinguish between two kinds of states of double vision. The first kind is illustrated by the scenario with the motorbike. In cases like this, I agree with Sarıhan that you have an

[15] Alternatively, Sarıhan could give a different argument for the claim that those states of double vision that are typically cited in favor of qualia are fully representational.

inaccurate visual experience of two objects. The second kind is illustrated by the forest example. In these cases, you enjoy an accurate visual experience. In the first scenario, you are not visually aware that you see double. In the second scenario, in contrast, it is phenomenally manifest to you that your experience involves double vision. Representationalists therefore have to identify those aspects of visual contents that make it manifest to you that you are seeing double without seeing two objects or scenes. Since a number of representationalists have provided plausible accounts of this, Boghossian and Velleman would have to exclude these options in order to make a convincing case for qualia.[16]

3. *The argument in favor of the confusion claim*

The confusion claim holds that Block and Kind confuse the fact that those states that are usually cited in support of qualia are automatically non-endorsed states with the fact that they are non-representational. In this section, I will address afterimages in order to argue that Block and Kind may not be confused about this.

Let me first quote the relevant passages from Block and Kind. Block writes: "[Afterimages] don't look as if they are really objects or as if they are really red. They ... look illusory" (Block 1996: 32; ellipsis in the original). And Kind writes: "But in none of these cases does it seem as if the afterimage represents something that is really there. When you close your eyes after looking at the bright light, for example, you don't take the lingering glow to be on the inner surface of your eyelids. When you see the red afterimage against a white wall, you don't take the redness to suggest the existence of a red dot on the page" (Kind 2008: 289). As you can see from these quotes, Block and Kind both describe afterimage experiences as illusory.

If I understand Sarıhan correctly, he analyzes the mistake in Block and Kind's reasoning as follows. Block and Kind describe afterimage experiences, using representational language. At the same time, they notice that the objects and properties referred to by these terms do not actually exist in the physical world. Consequently, they do not endorse the claim that these experiences represent material objects. But Block and Kind do not realize that the non-endorsement is a cognitive judgment and mistake it for a specific phenomenal feature of afterimage experiences, namely a feature that belongs exclusively to qualia. According to Sarıhan, Block and Kind thus confuse the fact that experiences of afterimages are non-endorsed states with the fact that they are non-representational. As I stated above, I agree with Sarıhan that Block and Kind make a mistake. But I do not think that they are confused in exactly the way described here. I will make two points in support of this.

[16] One such account can be found in Tye (2000). I also think that the account of perceptual content in Lycan (1995) can be extended to double vision.

My first point concerns the conclusion that Block and Kind confuse the fact that experiences of afterimages are non-endorsed states with the fact that they are non-representational. I have already argued in the previous section that states of double vision may not be accurately described as non-endorsed visual states. I think the same may be true for experiences of afterimages. A plausible way to interpret Block and Kind is as saying that afterimage experiences do not purport to represent material objects and their properties.[17] As Boghossian and Velleman put it, "afterimages are not seen as material objects, any more than, say ringing in one's ears is heard as real noise" (Boghossian and Velleman 1989: 87). Under certain circumstances, we may mistake afterimages for real objects. Schroer, for example, reports an afterimage caused by a lightbulb, which "he immediately took . . . to be a red beanbag" (2004: 543). Schroer may have had a temporary non-veridical experience of a red beanbag. However, it is plausible that genuine experiences of afterimages are not like that, that is, that they do not purport to represent material objects. If genuine afterimage experiences manifestly do not represent material objects, their contents can be accurate and dispose their subjects to endorse them at the cognitive level. This will happen if afterimage experiences represent their objects as what they are, namely as visual disturbances. I will elaborate on this in the next section. If this is correct, Block and Kind are not confused about the fact that afterimage experiences are non-endorsed visual states.

My second point concerns the representational language. In the passage quoted above, Block describes afterimages as looking as if they are not really red. According to Sarıhan, the most plausible way to interpret this is to take it to mean that something, some represented object, does not really look red. Similarly, Block describes afterimages as looking illusory. The most straightforward interpretation of this is to take it as saying that there is some brute feature of "unrealness" or "illusoriness" attached to the visually represented objects or properties, and this, so Sarıhan, is best understood as saying that the experience represents some object or some property as being unreal or illusory. Third, according to Sarıhan, it is also possible that Block means that something looks illusory in the sense that it is represented in a way that makes it unlikely that it exists. Similarly, Kind talks about a "lingering glow" and a "red dot," terms that are most plausibly interpreted as referring to representational contents. On all these interpretations, Block and Kind describe experiences of afterimages as representational states. But, so Sarıhan, this is inconsistent with their overall aim to argue in favor of qualia.

According to Sarıhan, one might interpret Block and Kind also in a different, more charitable, way. One might argue that they use representational language in order to refer to qualia, that is, to intrinsic,

[17] Phillips agrees with this. He writes, for example, "[A]fterimages simply do not appear to be material objects" (Phillips 2013: 425).

non-representational features of experiences. One could then say that Block and Kind hold that an afterimage is simply a complex of qualia. However, according to Sarıhan, this interpretation "has no appeal for people who, introspecting an afterimage experience, find no qualities other than those like colors that objects appear to have also in uncontroversially representational experiences, which makes it natural to think of afterimages as a special type of misrepresentation" (Sarıhan 2020: 25).

I find these observations about the use of representational language important. But I do not think that these observations clearly show that Block and Kind are confused about the intended referents of their terms. It seems plausible to me that they use representational language intentionally with the aim of describing intrinsic, non-representational aspects of experiences. Sarıhan correctly concludes that the phenomenological descriptions provided by Block and Kind may not convince those who do not find qualia when they introspect their own afterimage experiences. But the question at this point is whether Block and Kind are confused about the intended referents of these terms, and, as far as I can see, this may not be the case.

Sarıhan comes back to the more charitable interpretation of Block and Kind's use of representational language again towards the end of his paper. He considers the possibility of translating their descriptions into qualia theoretic terminology. He suggests, for example, that qualia theorists might use the term "red-quale" not in order to say that some object, a quale, is red, or that the quale represents redness, but rather to talk about the quale that we normally find in experiences that represent redness. Similarly, he suggests that qualia theorists might use the phrase "experiencing an afterimage" in a non-representational sense that would be similar to saying that we experience joy (Sarıhan 2020: 18). A successful translation of the entire representational language into qualia theoretic terminology would lend significant strength to Block and Kind's argument.

Sarıhan raises two problems for this interpretation, however. First, he argues that it is not plausible that we will be able to give a satisfactory translation of the complete description of these experiences. He points out, for example, that it is not very plausible to translate Kind's references to spatial relations, such as her claim that the afterimage is in front of the photographer's face, into qualia theoretic terminology. Second, he argues that there is no apparent reason for why we should translate the representational language into a qualia-theoretical vocabulary, since "we have a simpler analysis which treats these states as illusions of an automatically non-endorsed sort" (Sarıhan 2020, 30). I agree that it is implausible that qualia theorists will be able to give a satisfactory translation. I also agree that it is more plausible to take the descriptions at face value, that is, as a characterization of representational contents. But, if what I have said above is correct, these descriptions do not imply that afterimage experiences are non-endorsed states.

In my view, the most important problem with Block and Kind's argument is that they make use of an implicit assumption that is plausibly false. As I said above, we can interpret Block and Kind as saying that afterimage experiences do not represent colors and shapes as instantiated in material objects. Yet, the conclusion that these colors and shapes are therefore properties of experiences follows only if we also assume that viewers can experience visual properties only as instantiated either in material objects or in experiences.[18] In the next section, I will present a representationalist account of afterimage experiences that shows that there is an alternative. The representationalist can argue that afterimage experiences represent the colors and shapes as instantiated in pure visibilia.

Before moving on to the next section, I would like to summarize my argument so far. In section 1, I argued that there is a gap in the argument for the non-endorsement claim. What we would need is an argument for the claim that all states of double vision represent one thing, or one scene, as two. Absent such an argument, Block and Kind might still insist that the phenomenal character of genuine states of double vision outstrips their representational contents. In section 2, I made a parallel point about experiences of afterimages. Absent further argument, it is possible that the phenomenal character of genuine experiences of afterimages outstrips their representational content. In both sections, I also argued that these kinds of states might be accurate. If this is correct, Block and Kind do not confuse non-endorsed states with non-representational states. In order to fortify his arguments for the non-endorsement claim and the confusion claim, Sarıhan would therefore have to show that those states that are usually cited in support of qualia have inaccurate contents that conflict with the typical behavior of physical objects. If we assume, however, that these kinds of states are usually accurate, it follows that the representationalist is faced with a different task. The representationalist needs to give an account of the contents of these kinds of states that explains how they can be accurate. I will give such an account for experiences of afterimages in the next section.

4. *A representationalist account of afterimage experiences*

In this section, I will briefly sketch a representationalist account of afterimage experiences. My goal is to show that the implicit assumption underlying Block's and Kind's arguments in favor of qualia, that is, the assumption that viewers can experience visual properties only as instantiated either in material objects or in experiences, may not be correct. I will assume that Boghossian and Velleman are right in saying that genuine experiences of afterimages do not purport to rep-

[18] I believe that Sarıhan is aware of the fact that the arguments put forward by Block and Kind may presuppose this assumption.

resent material objects or properties instantiated in material objects. The main task for the representationalist then is to explain how it is possible for a viewer to be aware of colors and shapes that are instantiated neither in material objects nor in experiences. In the following, I propose that afterimage experiences represent a special kind of perceptual object, namely what Martin has called a "pure visible" (Martin 2012).[19] More specifically, I suggest that afterimage experiences represent these pure visibilia as visual disturbances.

My argument proceeds somewhat indirectly. I begin with an analysis of shadow experiences and then consider experiences of afterimages. I argue that both shadow experiences and experiences of afterimages represent pure visibilia.[20] But whereas experiences of shadows represent their objects as illumination phenomena, experiences of afterimages represent their objects as visual disturbances. I would like to emphasize that I do not understand the as-locution here in a conceptual sense. In order to see a pure visible as an illumination phenomenon or as a visual disturbance, you do not need to possess the relevant concepts. Rather, my claim is that your experience represents its content as an illumination phenomenon or a visual disturbance nonconceptually.

Let me begin with shadow experiences. We experience shadows in many different situations. For example, we see shadows when we watch a movie that is recorded on celluloid film and projected onto a screen. But when we watch movies, we enjoy visual experiences of ordinary three-dimensional objects – cars, trees, people, etc. These experiences are not genuine shadow experiences, that is, they are not experiences of shadows as shadows. In contrast, suppose that you are walking through a snow-covered forest on a sunny day. You see the blue shadows of the trees on the smooth white surface of the snow. This is a genuine shadow experience. In the following, I will talk only about these kinds of shadow experiences.

How do we describe such an experience phenomenologically? I think that the following features are uncontroversial. First, you are aware of shapes. These shapes have a determinate location – they are *on* the snow. These shapes lack a third dimension – they are genuinely two-dimensional.[21] Moreover, you do not experience these shapes as properties of the snow. Second, you are aware of the blue color of the shadows. But, again, you do not experience the blueness as the color of the snow – the snow looks white all over. You also do not experience the color as blue light that illuminates the snow. Experientially, the color belongs

[19] See also Phillips (2013).

[20] Phillips also argues that shadows are pure visibilia (Phillips 2018).

[21] One might object here that shadows are sometimes experienced as three-dimensional. This is true. For example, we sometimes experience silhouettes as three-dimensional. Moreover, moving shadows may imply three-dimensionality. On my account, these kinds of experiences are not genuine shadow experiences, but rather illusions, much like movies.

to the shadows. Genuine shadow experiences, we can say, represent colored two-dimensional shapes. Note that I use the term "color" here to include both chromatic and achromatic colors.

However, this does not exhaust the phenomenology of genuine shadow experiences. Even though you do not experience the blue color as blue light that illuminates the snow, you nevertheless experience it as an illumination phenomenon. The shadowed regions look to you like regions that receive less light than the unshadowed regions. This becomes clearer if we take the following two phenomenological facts into account. First, you see the snow as having the same color in the shadowed and the unshadowed regions. Second, you do not see the shadows as occluding the surface of the snow. The first fact makes clear that genuine shadow experiences involve color constancy. The second fact makes clear that this kind of color constancy is not a result of the visual system filling in the color behind a perceived material occluder, but rather an illumination phenomenon. As experienced illumination phenomena, shadows may improve or impede the perception of the visual properties of the shadowed regions.[22]

Martin has argued that we perceive shadows, and many other phenomena, as pure visibilia (Martin 2012).[23] He writes:

> [T]he visible world seems to contain both purely visual objects together with the concrete entities that we suppose are the medium-sized dry goods of the material world. We see lights, we see shadows, we see highlights, we see rainbows, we see the sky, and we can see mirror images or holograms; all of these things seem to be creatures solely of the visual world, nothing about them reveals how they would extend into physical space in aspects beyond those that we can detect visually. Alongside these purely visual phenomena, we see tables and chairs, rocks and sparrows, fires and hurricanes: entities which we can single out among visible objects, but which also have an existence and an impact well beyond the visible realm. (Martin 2012: 334)

If we assume that visual experiences are states with representational content, we can put this as follows. When you see a pure visible, your visual experience represents it as having only visible properties. Seeing a pure visible differs markedly from seeing a material object. When you see a material object, say a tree, your experience also represents it as having properties that are not visible. You see a tree as a three-dimensional solid object.

Now, according to the phenomenological description above, shadow experiences represent colored two-dimensional shapes that are located on surfaces. Since their visible nature is exhausted by their visible properties, they are pure visibilia in Martin's sense.[24] I further suggest-

[22] It is obvious that shadows often make it more difficult to clearly recognize the visible properties of objects. But sometimes, when the surrounding light is very bright, shadows can have the opposite effect.

[23] See also Phillips (2013, 2018).

[24] I would like to emphasize that I am not talking about the actual physical properties of shadows. Shadowed regions receive less light than the surrounding

ed that we perceive shadows also as areas that receive less light than the surrounding regions. I therefore think that it is correct to say that shadow experiences represent these pure visibilia as illumination phenomena. This claim needs to be understood in the right way. A shadow is the result of an object blocking the light from reaching the surface. But I do not claim that a shadow experience represents the causal origin of the shadow. I only claim that such an experience represents the shadow as a light phenomenon. I also want to emphasize again that this is a non-conceptual representational content.

I will now argue that afterimages also represent pure visibilia. But whereas experiences of shadows represent their objects as illumination phenomena, experiences of afterimages represent their objects as visual disturbances. To have a concrete case in mind, consider again Kind's scenario in which you experience an afterimage caused by a camera flash. After the flash goes off, you are aware of a patch (Kind says "dot") that has a complex shape and a color that is somewhere between blue and gray.[25]

How do we describe the experience of the patch phenomenologically? The patch has a rather complex phenomenology. I take it that the following phenomenal features are widely accepted:

1. The patch has a fuzzy border. This is similar to the borders of shadows. It is possible to create very sharp afterimages, but this requires very special circumstances. Typical afterimages have fuzzy borders.
2. The patch changes in appearance. When seen against a light background, the patch looks dark, and when seen against a dark background, the patch looks light. Similarly, the patch changes its appearance when you close your eyes.
3. The patch does not have a third dimension and it is not possible to look at the patch from different points of view.
4. The patch is not on the wall or at some determinate distance away from you. Phenomenologically, the patch seems to be floating in front of your eyes and this persists even if you close your eyes.
5. The patch moves with your eyes in much the same way in which floaters move. It is very difficult, or even impossible, to focus your eyes on the patch and follow its movement.

We can make additional observations, such as the fact that afterimages often seem self-luminous, that they often seem semitransparent,

regions and have complex physical properties. I suggest only that we perceive shadows as pure visibilia.

[25] I would also like to emphasize that experiences of afterimages are very varied, depending on the specific circumstances – the duration of exposure, wavelength and intensity of the light, conditions of observation (in darkness or light, with or without eye movements, etc.). In the following, I assume that the conclusions from Kind's example generalize to all genuine afterimage experiences.

that they can be refreshed through blinking, and that they disappear over time while changing color. But, for now, it suffices to focus on the features listed above.

These phenomenal features clearly support the claim that afterimage experiences represent pure visibilia. The patch has colors and shapes that are bound together. But these colors and shapes do not seem to be instantiated in a material object that has depth, can be seen from different points of view, moves independently of the eye, or has other properties that are characteristic of material objects, such as solidity. The representationalist can therefore plausibly claim that afterimage experiences represent pure visibilia.

Unfortunately, these phenomenal features alone do not suffice to support the second part of the claim, namely that afterimage experiences represent their objects as visual disturbances. One might plausibly argue that you can tell that the afterimage is a visual disturbance because of the way in which it moves and changes its appearance. But your ability to do so may be the result of a cognitive judgment and may thus not require special experiential contents. However, the features described above do not exhaust the phenomenology of typical afterimage experiences. Let me illustrate this again with Kind's example.

Suppose again that the camera flash goes off and that you see the patch floating before your eyes. If the afterimage is very intense, you will not be able to recognize any visible properties of objects that are located in the same direction in your visual field. If the afterimage is less intense, you may be able to recognize some, or even all, visible properties of these objects, but this will be more difficult. In both cases, you experience the patch as something that impairs your vision. However, like in the case of shadows, you do not experience the patch as a material object that occludes other objects located behind it, and, in contrast to shadows, you do not experience the patch as an illumination phenomenon. I therefore submit that you experience the patch as a visual disturbance.[26] Here too, I do not claim that the afterimage experience represents the causal origin of the afterimage. The experience does not represent that the afterimage is a result of certain photochemical processes taking place in your retinal cells, for example. Rather, the experience represents the afterimage just as a visual disturbance.

This concludes my brief sketch of a representationalist account of afterimage experiences. In my view, genuine afterimage experiences represent pure visibilia (e.g., the patches caused by camera flashes) as visual disturbances. If this is correct, it is plausible that the phenomenology of an afterimage experience is exhausted by its representational content. Moreover, since afterimages are in fact visual disturbances, genuine afterimage experiences can be accurate. They are not illusory

[26] This is also indicated by the fact that you do not try to improve your vision by moving the patch out of the way. Rather, you are tempted to make the afterimage go away by blinking or rubbing your eye.

as Sarıhan suggested. This account of afterimage experiences has two consequences that are important for the argument in this paper. First, the fact that afterimage experiences represent afterimages as visual disturbances accommodates Boghossian and Velleman's phenomenological observation that they do not purport to represent material objects. Second, the account implies that Block and Kind mistake a certain kind of representational content, namely pure visibilia, for qualia.

5. *Conclusion*

In sections 1 and 2, I argued that the arguments in favor of the non-endorsement claim and the confusion claim contain a gap. In order to fill the gap, Sarıhan would have to show that states of double vision and experiences of afterimages have non-veridical contents. However, if it is not possible to give such an argument, the representationalist needs to give an account of the contents of these kinds of states that explains how they can be accurate. In section 3, I sketched such an account for afterimage experiences, suggesting that they represent pure visibilia. If this account is on the right track, it follows that the implicit assumption in Block's and Kind's arguments, that is, the assumption that visual states can represent colors and shapes only as instantiated either in material objects or in experiences, is false. It then also follows that Block and Kind do not confuse the fact that afterimage experiences are non-endorsed mental states with the fact that they are non-representational. Rather, they fail to see that we can experience colors and shapes as instantiated in pure visibilia.[27]

References

Block, N. 1996. "Mental paint and mental latex." In E. Villanueva (ed.). *Philosophical Issues*, vol. 7: Perception. Ascadero: Ridgeview: 18–49.

Boghossian, P. and Velleman, D. 1989. "Color as secondary quality." *Mind* 98: 81–103.

Byrne, A. 2001. "Intentionalism defended." *Philosophical Review* 110 (2): 199–240.

Dretske, F. 1995. *Naturalizing the mind*. Cambridge: The MIT Press.

Kind, A. 2008. "How to believe in qualia." In E. Wright (ed.). *The Case for Qualia*. Cambridge: The MIT Press: 285–298.

Lycan, W. G. 1996. *Consciousness and experience*. Cambridge: MIT Press.

Martin, M. G. F. 2012. "Sounds and images." *British Journal of Aesthetics* 52 (4): 331–351.

Masrour, F. 2013. "Phenomenal objectivity and phenomenal intentionality: In defense of a Kantian account." In U. Kriegel (ed.). *Phenomenal Intentionality*. Oxford: Oxford University Press: 116–136.

Peacocke, C. 1983. *Sense and content*. Oxford: Oxford University Press.

[27] I would like to thank Chris Cuomo, Ed Halper, Aaron Meskin, Beth Preston, and Sarah Wright for their very helpful comments on this paper.

Phillips, I. 2013. "Afterimages and sensation." *Philosophy and Phenomenological Research* 87 (2): 417–453.

Phillips, I. 2018. "No more than meets the eye: Shadows as pure visibilia." In T. Crowther and C. M. Cumhaill (eds.). *Perceptual Ephemera*. Oxford: Oxford University Press: 172–191.

Sarıhan, I. 2020. "Double vision, phosphenes and afterimages: Non-endorsed representations rather than non-representational qualia." *European Journal of Analytic Philosophy* 16 (1): 5–32.

Shroer, R. 2004. "Environmental representationalists on afterimages and phosphenes: Putting the best foot forward." *Southern Journal of Philosophy* 42 (4): 531–546.

Tye, M. 1995. *Ten problems of consciousness: A representational theory of the phenomenal mind*. Cambridge: The MIT Press.

Tye, M. 2000. *Color, consciousness, and content*. Cambridge: The MIT Press.

Croatian Journal of Philosophy
Vol. XX, No. 60, 2020

The Limits of Expertism

NENAD MIŠČEVIĆ
University of Maribor, Maribor, Slovenia
Central European University, Budapest, Hungary

Snježana Prijić-Samaržija's book discusses the epistemic grounding of democracy, stressing the epistemic role of experts in her political-epistemological favorite, the project of "reliability democracy". Her proposal, inspired by Christiano, lets citizens play an important role in setting the aims, whereas experts deliberate about means of reaching them. I argue that it is not easy to reach consensus about goals and values. What is needed is democratic deliberation in deciding, encompassing both experts and laypersons. We should retain the duality of less ideal deliberation in real world and of hypothetical contractualist deliberation, within moral-political thought-experiments, in the tradition of Habermas and Scanlon in the ideal theory. I leave it open whether our author might ultimately agree with this picture of reliability democracy.

Keywords: Democracy, epistemology, deliberation, experts.

1. *An important defense of hybrid virtues*

In the present paper I want to discuss the (2018) book by Snježana Prijić-Samaržija entitled *Democracy and Truth: The Conflict between Political and Epistemic Virtues*.[1] I find it to be a very important book, and a fine continuation of Snježana's work on social epistemology (I shall be using her first name, "Snježana", to refer to her, since we have been friends for decades.) The book is bringing together epistemic virtues and political goals, systematizing and presenting a lot of recent literature and offering an original view on the role on experts in public deliberation. (It should be translated into Croatian for home readers!). I am happy to be

[1] The paper was written for the issue of the *Croatian Journal of Philosophy* on political philosophy. Unfortunately, I made a technical mistake, and did not submit it for the issue. Snježana-Prijić did write some lines of reply (Prijić Samaržija 2020a), using the manuscript, and the reply appeared in the issue to which I did not contribute my paper on time. I thank her a lot, and I apologize for the mistake I made. So, the paper is appearing in the present volume.

able to discuss it with the author; we already had a round of discussion in Maribor, and it was extremely useful, to me, a least.

Here is then the preview. The first section gives a brief sketch of the views proposed in the book, concluding with one of the central issues from it, namely the role and the choice of experts. I argue that Snježana's proposal is opened to a difficult dilemma, worthy of addressing in detail. The second section contains a brief sketch of my contractualist proposal concerning the choice of experts and the nature of public deliberation. I leave it open whether the proposal is compatible with Snježana's views; this is in fact my main question for her.

In recent presentations, for instance in Maribor and Belgrade, both in 2019, she has placed her proposal in the political context of present "crisis of democracy". She listed the well-known symptoms, most of which we all recognize:

> Radicalization, terrorism, fundamentalism, souverenism, xenophobia, nationalism, chauvinism...
> Trump, Brexit, Orban… different kinds of populism.
> Pseudo-science: anti-vaxxers, creationism, climate changes deniers, etc.
> Too much democracy?
> Too little (real) democracy?

We shall return to this context at the end of the first section and stay within it till the end.

The book itself begins with a telling quote by Emma Goldman, connecting an epistemic state – ignorance with a political dimension of violence: *"The most violent element in society is ignorance."* (1910: 3). Ignorance comprises "reliance on stereotypes and prejudices, evident disregard for rational and responsible decision-making (…)", and an obvious lack of awareness about cognitive and evidential limitations, epistemic egoism and the like (Prijić-Samaržija 2018: 11). Snježana then notes: "Yet, while violence is consistently faced with unanimous condemnation, the ignorance from which it stems remains exempt from any kind of direct scientific scrutiny" (2018: 11).

Snježana then gives an excellent overview of the political epistemology in the last seven or eight decades, from Rawls to present-day thinkers. She criticizes political instrumentalism (which she ascribes to political philosophers from Rawls to epistemic proceduralism, consensualism, but also to postmodernism (authors like F. Peter, D. Estlund, P. Kitcher, and M. Fricker), i.e. reduction of epistemic virtues to political virtues. She is equally critical of epistemological instrumentalism, the view which favors reduction of political virtues to epistemic virtues (of truth) (e.g. epistemologists in standard analytical epistemology, from Goldman's very moderate stance to Neven Sesardić's radical versions. She talks of the anti – democratic character of both political and epistemological instrumentalism (mostly in chapter Four of the book). Her sympathies lie with what she calls "hybrid standpoint", favoring

harmonization of political and epistemic virtues instead of reduction.

So she criticizes both elitism (epistemic instrumentalism that prefers truth to justice) and egalitarianism (political instrumentalism, embodied both in epistemic proceduralism and consensualism); this is done in section 4 of chapter Two. Her preferences are with the hybrid standpoint, involving contextualism and localism. The main area of application of the standpoint is justification of democracy (Chapters Three to the concluding chapter Six). I think her discussion of hybrid virtues is highly relevant: it is an important defence of a plausible standpoint in political epistemology. Let me add that in my view this hybrid combination would work also on the critical side, for the purpose of criticizing given social arrangements for being both epistemically and politically bad; this would be an interesting line in critical social epistemology.

Let me note her attitude to Habermasian project of deliberative democracy; I shall express my partial disagreement in a moment. Habermas started his project of contractualist justification of politics by appealing to deliberation in idealized circumstances. But then in his work *Between Facts and Norms* (1992) he connected it to normal, non-ideal "deliberative politics". To quote a standard reading of his final result, building on his notion of discourse, Habermas understands deliberative politics to refer to institutionalized discursive procedures of will formation and decision making within constitutional political systems (Deitelhoff 2018: 528).

Snježana notes that she accepts his consensualism, but rejects his picture of the procedure that is to lead to consensus:

> The fundamental problem of Habermas' consensualism, consensualist theory of making fair decisions, or his consensualist theory of justice, is its dubious application in conditions that are less than ideal. Achieving a rational consensus in idealized communicational circumstances – implying a sincere, tolerant, argument-based and informed, institutionally well organized and guided debate – is certainly an attractive and highly acceptable goal. However, any sub-ideal moral and epistemic circumstances render the concept of a 'rational' consensus entirely unclear. (2018: 168)

Finally, in chapter Six, Snježana opts for "reliability democracy" characterized by fulfilling five veritistic criteria that should guarantee the epistemic quality of a procedure. Here is her list, inspired by Goldman: (i) *reliability*, or the ratio of true and false decisions generated by this procedure; (ii) *power*, or total productivity in producing beliefs; (iii) *fecundity*, or the capacity of a social practice or institutions to solve the problems of interested citizens; (iv) *speed* or the time required (…); (v) efficacy; or the cost of achieving goals (2018: 202)

As can be guessed from what I said until now, I agree with a lot of what Snježana says. Most importantly, about truth being the goal of deliberation.

One of the main topics of the book is the positive role of experts in the social and political cognition, as typical for reliability democ-

racy. She borrows some ideas from Thomas Christiano on the division of epistemic labor, where lay-persons choose the aims of society, and experts decide about means. Here is Snježana's summary of her view from her recent paper discussing her book:

> Although experts are conventionally excluded from democratic procedures out of fear of epistocracy, or of undemocratic elite privilege, I hold that the exclusion of experts is a conscious sacrifice of epistemic quality, and, consequently, of the best democratic decisions. If epistemic justification is required for justifying democracy, then excluding experts is just as undemocratic as excluding citizens. (2020: 59)

Finally, the relevant chapter of the book, culminates with the following statement:

> So, in contrast to the reliabilist concept of externalism in which decisions are epistemically and democratically justified as long as there are reliable democratic mechanisms that produce truth-sensitive decisions, I would like to stress a certain need for more participation or for better epistemic and democratic access of citizens and policy makers to decisions. More precisely, while the responsibility of a reliability democrat would be to insure a reliable democratic procedure, the internalist approach, in whose favor I am arguing, stresses that it is necessary for citizens and policy-makers to understand why it is rational to rely on expertise and reliable democratic procedures and why it is rational to trust to these very procedures. Even if citizens and policy-makers cannot have full understanding or possess the total body of evidence to appraise the whole content of the experts' stances, their reliance or trust would be epistemically justified as long as they have enough evidence about the reliability of the procedures through which experts make their decisions. (2020: 216)[2]

So what is required is the division of roles: consensus is fine for ideal theory, for the non-ideal one we need experts (2018: 170). How are the experts to be chosen? Snježana proposes a consensual choice of experts. But what should the consensus be like? In the rest of this paper, I shall concentrate on varieties of consensual choice, and point to the deliberative procedure as a possible important context of such a choice.

So, how does the choice proceed? Here, there are different readings that are possible, and I shall concentrate on the two of them.

Reading one: the lay-population *as a whole* chooses the experts.

Reading two: various groups within lay population choose each their own expert(s).

[2] And she continues about the evidence of reliability of experts' procedures:

> For instance, that could consist of evidence about the experts' moral and epistemic characters (or the reputation of the institutions), evidence about the contextual (conversational) circumstances that prevent deception, lying and incompetence or support trustworthiness, or even proof of the presence of Christiano's truth sensitive mechanisms such as solidarity, overlapping understanding, competition and sanctions. More precisely, the democratic division of epistemic labor needs to embrace more epistemic agency on the side of citizens: they should have an active role in assessing which particular experts deserve trust and whether reliable mechanisms truly preserve the experts' trustworthiness. (2020: 216)

Snježana's explicit formulations seem to go more in the direction of the first reading. She talks about consensus preceding the choice; but the choice is consensual, and grounded in the will of the whole lay population. She appeals to Rawls' idea of "overlapping consensus" preceding important decisions, and not following from them. In her proposal it is a consensus generated by "public debate and conciliation of all interested participants" (2018: 192); she sees her view as "a certain amalgam of Rawls's and Habermas' proposals" (2018: 192)

If the first option is her choice, then we have a point of disagreement: I would opt for the second option. Let me develop the matters using Snježana's characterization of our political context, namely crisis of democracy. It encompasses, as we noted at the beginning of the paper, phenomena like

> Radicalization, terrorism, fundamentalism, souverenism, xenophobia, nationalism, chauvinism...
> Trump, Brexit, Orban…different kinds of populism
> Pseudo-science: anti-vaxxers, creationism, climate changes deniers, etc.

We, the truth-respecting theoreticians, have two options.

> Option one, let experts have a final word both on goals-values and on means to reach them. Option two, citizens contribute their values, and decide about aims.

But, how do we arrive at a consensus about who is the expert on such situation(s), consensus that is to precede political deliberation?

Start with option one: experts alone are to decide. But then, who are the typical experts for goals and value? The usual criteria don't fare well with this question. Look at the list of criteria i.e. summary of the traits "usually associated with expertise" recently put together by Carlo Martini in his very informative (2020) overview on social epistemology of expertize. He notes that "[E]xperts back their judgments with arguments, and present evidence in support of their opinions (Martini 2020: 118). He calls this trait "objectivity". Further, that "/E/xperts have credentials, usually as a track-record of their experience in the relevant field. And also that "experts give judgments within their field of competence, they do not judge without qualifications on matters that are not in their field of expertise. (Martini calls it "Pertinence (domain)" (Martini 2020: 118).[3] Can we really imagine laypersons from our surrounding reaching consensus about who is the expert on, say, morality of abortion? Perhaps doctors, but not priests, some average women might claim. The pregnant women themselves and no one else,

[3] Further traits listed are social acclamation, unbiasedness. (i.e. "possession of content-knowledge; that is, information specific to the field in which they possess expertise." and meta-knowledge. ("Experts know how much they know and how much they do not know.") (Martini 2020: 118). Finally, they exhibit consistency and are also able "to discriminate between very similar but not completely equivalent cases" (Martini 2020: 118).

the more feministically oriented women would claim. Priests above all, our religious compatriots would say. No way, simple aggregation of lay votes would give nothing useful. On option two, citizens contribute their values, and decide about aims. Thomas Christiano proposed it and Snježana seems to agree with him.

> citizens rule over the society by choosing the aims of the society and experts, along with the rest of the system, are charged with the tasks of implementing these aims with the help of their specialized knowledge. (Christiano 2012: 51)

We need experts on facts and this is obvious and quite trivial. But what about moral issues? We know that their bad mistakes are typical. Or at least, the disagreements are not easy to avoid. For the moment let us stay with Snježana's list and take xenophobia as a typical anti-democratic stance. Here is an example of a contrast in attitudes, characteristic for the presen decade.

Many citizens of Serbia accept that their country should be hospitable (to non-enemies). Citizens and experts, say in Belgrade, further agree that refugees are not enemies. Therefore, Serbia is not being oppressive towards refugees that come from Macedonia, and want to continue towards Germany.

In contrast, many citizens of Croatia favor the value of safety. Citizens and experts agree that refugees are the threat to safety. Therefore, Croatia becomes oppressive towards refugees coming from Bosnia. Our former president said a year ago: you need some brutality to deal with them (and police brutality in this year, 2020, at the border with Bosnia, has become truly unbearable). So, both options look problematic. Looks like a dilemma for Snježana. What can be done? Snježana has interesting constraints on deliberation:

> Our ongoing discussion about the epistemic justification of deliberative democracy has outlined several key conditions that should be incorporated into democratic procedures in order to ameliorate their reliability: (i) education – public discussions and exchanges of reasons should be based on educational and informative content that improves the participants' ability of conscientiously deliberate about various topics (ii) diversity – public debates should include citizens and experts with different perspectives who come from different communities and institutions, thus expanding the available pool of evidence (iii) non-egoism (inclusiveness, fairness, pluralism) – participants in public debate, both citizens and experts, must be aware of the cognitive constraints imposed by their presuppositions, worldviews and value systems, come to terms with their capacity to understand certain topics and maintain a disposition of openness towards different perspectives (iv) institutional organization – public discussions and decision-making processes must be initiated, monitored and guided by relevant institutional procedures that guarantee adherence to prerequisites (i) – (iii). (Prijić-Samaržija 2018)

Some of these might help, in particular diversity and non-egoism. However, they are extremely demanding on ordinary citizens. To stay with abortion example, I find it very difficult to take the pro-abortionist per-

spective in my ordinary citizen role. And I don't think the difficulty comes from my egoism! A non-philosopher citizen, say a woman with firmly anti-abortionist stance, will find it even more difficult**.**

In brief, there is *no neutral, consensual* way to choose experts. Any final decision has to involve experts as well; if we want a kind of reliability democracy, we shall have to trust the interaction between the two, and respect differences in view where experts with a certain attitude are aligned with laypersons sharing the attitude. But then, the deliberation has to accompany the choice of experts, and cannot precede it.

2. *Could Deliberation Nevertheless Help?*

Snježana herself noted, and we quoted her, that consensus is fine for ideal theory, whereas for the non-ideal one we need experts (2018: 10). She then assumes that "any sub-ideal moral and epistemic circumstances render the concept of 'rational' consensus entirely unclear" (2018: 168). But why not look for parallels between the possibilities of ideal and of non-ideal theory; the two together might offer some hope to avoid the dilemma. Namely, there are parallels between ideal and sub-ideal constellations that enable us to do the following: first, we can project the notion of 'rationality' downwards, from the ideal to the sub-ideal: a sub-ideal constellation is 'rational' if it is sufficiently close to its ideal counterpart. Second, we can go into opposite direction, and ask about a sub-ideal constellation if it approximates its ideal counterpart.

Take the abortion example (we shall look at the other, xenophobia example few lines below). In non-ideal situation we shall have groups of similarly minded citizens, encompassing both expert and lay-persons. Religious Catholic Croats will normally side with local priests, feminist lay-persons will side with feminist lawyers and social psychologists, and so on. And the groups, each containing experts and similarly minded lay-persons will enter "network of discourses and bargaining processes" as Habermas calls them (1996: 320). We can then project the division right into the Ideal formulation: idealized representatives of each group would debate the relevant issue between them, and the debate, in favorite cases, will produce truth or some similar epistemic justification.

A similar picture is offered by Scanlon:

> The central component of individual morality as I understand it – what I call the morality of what we owe to each other – is something we have reason to care about because we have reason to care about our relations with others in which justifiability of this kind plays an important role. There is, I believe a corresponding version of the morality of institutions, consisting of standards that institutions must meet if they are to be justifiable to those to whom they claim to apply. (2016: 20)

My proposal for answering Snježana's doubt is that we retain the duality of less ideal deliberation in real world and of hypothetical contrac-

tualist moral-political thought- experiments, in the tradition of Habermas and Scanlon in the ideal theory.[4]

Let me stay with non-ideal approximation for the moment. Take the relevant division of a given society, say the class, gender, ethnic, religion or any other and imagine the following:

Let each group have its own experts: workers get the class-conscious intellectuals, entrepreneurs get Nozickian neo-liberal experts to articulate their needs, and so on.

Think of a representative of each group: she is representing both its experts and its popular basis, and then imagine them interacting, mostly by debating. They search for principles that

"could not reasonably be rejected, by people who were moved to find principles for the general regulation of behavior that others, similarly motivated, could not reasonably reject.", as Scanlon would put it (1998: 4). In the process the proposals are made by parties, each of which include both the ordinary citizens and the similarly oriented experts. Might this solve our dilemma? Or is it just a phantasy?Let me offer an extended example in favor of the optimistic answer. Start in a realistic spirit, with actual, non-ideal conditions. What do the efforts of increasing public rationality normally look like in such conditions? Start with Snježana's examples of political irrationality, say xenophobia, and consider anti-refugee xenophobia spreading from Greece to Germany and all the way to Scandinavia.

Some efforts against it have been and hopefully more will be done within civil society, some on the higher, administrative level. Orthogonally to this division, there is also the width division, the national vs. the global. Combine the two, and you will get four pigeon-holes, which are drawn three lines below. There are well-known examples for each, some listed within the table:

[4] Here we agree with Robert E. Goodin and Kai Spiekermann:

> Theorists of deliberative democracy are undoubtedly correct in thinking that it would be better—in epistemic as well as in many other respects—if interpersonal interactions were governed by the high standards approaching Habermas's 'ideal speech situation'. A raft of small-scale experiments trying to do just that show that, after formal deliberations in which moderators enforce such rules, people's opinions are different in all sorts of ways that would presumably make them more competent voters. (We are skipping here the footnotes of the authors-NM) Although those are highly stylized deliberative settings, they are not without real-world political relevance. Many of those same standards are written into manuals of parliamentary practice, after a fashion. Even where they are not, they typically figure at least in manuals of good manners. Of course, both sets of instructions contained are often honoured in the breach. Still, it may not be beyond hope that those ideals might be approximated in the real world, at least in certain settings. Whether those experiences and experiments can be scaled up to the societywide level is an open question. Our point here is simply that they do not need to be. The epistemic benefits that come from interpersonal interaction (two-way, or even just one-way) do not completely depend upon realization of those higher deliberative democratic ideals as a society-wide exercise—epistemically better though it would no doubt be, if that were realized. (2018: 134)

	Narrow	*Wide-global*
Civil	Croatian and Slovenian NGOs acting in favor of refugees	Red Cross helping
Administrative	Merkel inviting refugees to Germany in 2014.	UNHCR activities. Marakesh migration pact

However, this is only the beginning. The next important division concerns constructive vs. critical activities. The former are omnipresent, the later are more characteristic of intellectual and journalistic area.

Next, look at the motivation. We can follow Habermasian inspiration, and distinguish between more prudential, more moral, and more legal-cum-political motivation. Merkel's motivation for inviting refugees had a prudential component, German's need of skilled workers, but it might have also been prompted by her moral attitudes. Marakesh pact is formulated in legal-cum-political terms, but one can perhaps hypothesize some moral stance in the background.

Some actions will be exclusionary. On the positive side here is a bottom-up critical effort – belling-the-cat: warn common people against the lies and injustices directed against the refuges by establishment agents. (The Bellingcat enterprise did similar actions supporting minorities and the like.)

What about experts and laypeople in this context? Some activities will demand experts, for instance critical journalist ones. Some demand collaboration between specialized experts and run-of-the-mill politicians. Others demand wide lay, non-expert engagement in order to succeed.

The hope is that there will be positive, virtuous circles along the line: experts will become more politically aware through their interaction with lay-people and vice versa. Next, that there will be positive, virtuous circles both in the top-down and in the bottom-up directions. And, most important for our purposes, there will be interaction between the pro-groups and anti-groups, which *might help towards more rationality* in political life.

Take any of the items on Snježana's list: radicalization, terrorism, fundamentalism, souverenism, nationalism, chauvinism, then Trump, Brexit, Orban…different kinds of populism, and finally pseudo-science: anti-vaxxers, creationism, climate changes deniers, etc.

For each of them we can find examples of interactions similar to the ones we listed for the case of refugees.

Call all these interactions "deliberative interactions". The hope is that they ultimately embody the requirements of deliberative democracy For instance, for fighting the climate change denying discussed in the context of Habermas and deliberative democracy see Emilie Prattico. Here is a quote:

> Once the intricacy of the various modalities of reason that are at play in political problems becomes visible and, importantly, contestable through discourse, the Frankfurt School's insight that our fate is tied to that of nature

> descends to street-level, so to speak. Indeed, through deliberation we reach a position whence we can see, not only that ecological questions are also questions of justice, but that it is by seeking political resolutions to such problems through discourse that we might reasonably hope to solve them at all. There is no 'getting it right' without 'getting it fair', too. (Prattico 2019)

Scanlon comes very close to the same general stance; in his paper on institutional morality he points to parallels between matters of justice and matters of "public policy" (2016: 18). Once Habermas and Scanlon are in play, we can pass to the other side of the matters. We have looked at debates in actual, non-ideal circumstances. But Habermas and Scanlon primarily offer an ideal-theoretic version of debating, which can be applied to the same issues. They ask us to imagine a social-contract scenario in which participants debate the burning issues; the new element is that the debating situation is somewhat idealized. We are asked to imagine the participants being more rational and fully informed. The propositions that such participants could agree upon would be written down in this hypothetical contract. (Habermas discusses the rationale for idealization, and a kind of parallel between the idealized and the real context in his (2001) dialogue with Thomas McCarthy).[5]

Imagine, to stay with our standing example, an idealized debate concerning refugees. In the debate refugees should have a representatives, and potential host countries some representative as well, say one or two for, the rich and the middle. Now the representative, call them "the Rich" and "the Middle" are supposed to have full information; most importantly to know that, for example, Muslim refugees present no cultural threat to their countries. Next, they are supposed to be fully rational, capable of deriving consequences from initial premises.

We can imagine that ultimately the sides, the Refugees, the Rich and the Middle (representatives) might agree about a moderate right for the Refugees, for instance that the right of asylum is a defeasible human right (as Kieran Oberman proposes in his 2016 paper).

What would be the non-ideal counterpart of such agreement in our Croatian-Serbian example? One difficult to implement but easy to imagine solution would involve first, the coming of both countries under the same political umbrella, most simply, Serbia entering the EU. The next demand would be unifying the relevant legal arrangements. In the optimistic case, Croatian police would start following the same tolerant practice that Serbian police has been implementing in last five years in favor of refugees.

The hope is that there is *a significant parallel between the ideal and the non-ideal version of the deliberative ideal*. If the hope is realistic, as we hope it is, then Snježana's doubts about the validity of rational deliberation in non-ideal situations can be laid to rest.

[5] He concludes his discussion with the following telling statement:

> …with idealizations we explain from a participant's perspective the operations that actors must accomplish in their actual performance of certain everyday practices, namely those we describe as communicative action and rational discourse. (McCarthy 2001: 37)

3. *Conclusion*

Snježana's book is a highly original work, covering several fields within social epistemology, from the role of truth in social-political justification to the relatively detailed questions concerning epistemic legitimacy of democracy. Her final proposal of a version of reliability democracy, featuring important epistemic-political role of experts, is promising and applicable to the actual crisis of democracy that we all witness.

The present paper has concentrated on the role of deliberation and the choice of experts. Snježana seems to favor placing democratic deliberation at the beginning of the building of consensus, with the role of designating experts, who then make further, epistemically solid decisions concerning the problems at hand. She describes her project as a synthesis of Habermas and Rawls, but with firm reservations about Habermas' contractualist project.

In the paper I ask the natural question: why not think of deliberation in contractualist tradition, proposed by Habermas and Scanlon and occasionally either praised or criticized by Snježana? I propose that what Snježana rightly sees as "the fundamental problem of Habermas' consensualism", i.e. "its dubious application in condition that are less than ideal" (2018: 168) can be hopefully solved starting from the assumption that there is a parallel between the ideal and the non-ideal version of deliberative ideal. Authors like Gutmann and Thompson take deliberative democracy to be "an aspirational ideal" (2004: 37); the task is then to look at real-world, non-ideal approximations to the ideal, instead of separating the two by an iron fence. The issue of expertize then becomes part of the characterization of process of deliberation, which starts with expert-laypersons grouping around particular stances on particular issues, continues with debate and deliberation performed by disagreeing groups, led by their respective expert intellectual leaders, and hopefully converging towards a rational consensus.

So, let me conclude with a question for Snježana: How far would you agree with my picture? Or would you disagree completely and if yes, why?

References

Christiano, T. 2012. "Rational deliberation among experts and citizens." In J. Parkinson and J. Mansbridge (eds.). *Deliberative systems: Deliberative democracy at the large scale,* Cambridge: Cambridge University Press: 27–51.

Deitelhoff, N. 2018. "Deliberation." In H. Brunkhorst, R. Kreide, and C. Lafont (eds.). *The Habermas handbook*. New York: Columbia University Press: 528–532.

Goldman, E. 1910. *Anarchism what it really stands for,* available at https://www.panarchy.org/indexes/panarchy.html

Goodin, R. E. and Spiekermann, K. 2018. *An Epistemic Theory of Democracy*. Oxford: Oxford University Press.

Gutmann, A. and Thompson, D. 2004. *Why Deliberative Democracy?* Princeton: Princeton University Press.

Habermas, J. 1984. *Theory of Communicative Action, Volume 1: Reason and the Rationalization of Society*. Boston: Beacon Press.

Habermas, J. 1987. *Theory of Communicative Action, Volume 2: Lifeworld and System: a Critique of Functionalist Reason*. Boston: Beacon Press.

Habermas, J. 1988. *Theory and Practice.* Boston: Beacon Press.

Habermas, J. 1994. *Justification and Application: Remarks on Discourse Ethics.* Cambridge: The MIT Press.

Habermas, J. 1994. "Three Normative Models of Democracy." *Constellations* 11: 1-10.

Habermas, J. 1996. *Between Facts and Norms.* Cambridge: The MIT Press.

Habermas, J. 2000. *On the Pragmatics of Communication*. Cambridge: The MIT Press.

Habermas, J. 2001. "From Kant's Ideas of 'Pure' Reason to the Idealizing Presuppositions of Communicative Action: Reflection on the 'Detranscendentalized' Use of Reason." In J. Bohman and W. Rehg (eds.). *Pluralism and the Pragmatic Turn: The Transformation of CriticalTheory, Essays in Honor of Thomas McCarthy.* Cambridge: The MIT Press: 11–40.

Habermas, J. 2003. *Truth and Justification.* Cambridge: The MIT Press.

Martini, C. 2020. "The Epistemology of Expertise." In M. Fricker, P. J. Graham, D. Henderson, and N. J. L. L. Pedersen (eds.). *The Routledge Handbook of Social Epistemology*. New York and London: Routledge: 11511–122.

Oberman, K. 2016. "Immigration as a Human Right in Migration." In S. Fine and L. Ypi (eds.). *Political Theory: The Ethics of Movement and Membership*. Oxford: Oxford University Press: 321–56.

Prattico, E. 2019. "How-can-Habermas-help-us-think-about-climate-change." *Aeon*, Dec 18,

Prijić-Samaržija, S. 2018. *Democracy and Truth: The Conflict between Political and Epistemic Virtues*. Monfalcone: Mimesis.

Prijić Samaržija, S. 2020a. "The epistemology of democracy: the epistemic virtues of democracy." *Philosophy and Society* 31 (1): 56–68.

Prijić Samaržija. S. 2020b. "The Epistemic Justification of Democracy." *Croatian Journal of Philosophy* 59: 183–198.

Richardson, H. S. 2002. *Democratic Autonomy: Public Reasoning About the Ends of Policy.* Oxford: Oxford University Press.

Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge: Harvard University Press.

Scanlon, T. M. 2008. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge: Harvard University Press.

Scanlon, T. M. 2016. "Individual Morality and the Morality of Institutions" *Filozofija i društvo* 17 (1): 20–37.

Scanlon, T. M. 2018. *Why Does Inequality Matter?* Oxford: Oxford University Press.

Croatian Journal of Philosophy
Vol. XX, No. 60, 2020

Two Concepts of the Epistemic Value of Public Deliberation

JOHN B. MIN
College of Southern Nevada, North Las Vegas, USA

Epistemic justification is necessary for deliberative democracy, yet there is a question about what we mean by the concept of epistemic values of public deliberation. According to one reading, the epistemic value of public deliberation implies a procedure's ability to achieve a correct outcome, as judged by a procedure-independent standard of correctness. As I shall show in this paper, however, there is another reading of the "epistemic" value of public deliberation extant in the literature: Epistemic values are constitutive of a deliberative process as an exchange of reasons. If the distinction between two concepts of epistemic values of public deliberation holds, then we can re-conceptualize the relationship between procedural fairness, epistemic values, and legitimacy. Thus, a concept of legitimacy that combines procedural fairness and a procedure-independent standard of correctness on the one hand, versus one that combines procedural fairness and the constitutive epistemic value of deliberation on the other hand.

Keywords: Deliberative democracy, epistemic democracy, democratic legitimacy, epistemic proceduralism.

Legitimacy is one of the central normative concepts for deliberative democracy because decisions must be mutually justifiable to all.[1] The legitimacy of decisions is a function of procedural and epistemic dimensions.[2] While there is a tendency to regard the procedural and epis-

[1] The literature in deliberative democracy is extensive. For good overviews, see Bohman and Rehg (1997), Chambers (2003), Marti and Besson (2006), Thompson (2008), Mansbridge et. al. (2012), Owen and Smith (2015). For statements about the centrality of legitimacy in deliberative democracy, see Cohen (1997a), Habermas (1996), Chambers (2003), Thompson (2008), Dryzek (2010), and Mansbridge et. al. (2012), among others.

[2] Let me clarify why I am using the "procedural and epistemic" distinction here. I am referring to procedural and epistemic justifications of democracy. For example, Estlund (2008) argues against pure proceduralist theories of democracy

temic dimensions as mutually exclusive categories of justification, recent literature suggests novel ways to combine those categories.[3] This essay addresses one strand of the multifaceted debates over the relationship between epistemology and democracy[4], namely the relation-

and pure epistemic theories of democracy, ultimately arriving at his favored theory of "epistemic proceduralism." Another example is Peter (2008) who argues against pure proceduralist views and instrumental theories of epistemic democracy such as Dewey's pragmatic view, ultimately arriving at pure epistemic proceduralism. Thus, I do not believe that the procedural and epistemic distinction vis-à-vis democratic legitimacy is unwarranted. However, I acknowledge that the usual and more familiar distinction is between procedural and substantive on the one hand and moral and epistemic on the other hand. The procedural and substantive distinction on the one hand is commonplace in political and legal philosophy. In the deliberative democracy literature, Cohen's (1997b) classic paper, "Procedure and Substance in Deliberative Democracy" is about that. Also, Gutmann and Thompson (1996) make a distinction between procedure and substance, where they give three procedural and three substantive conditions of deliberative democracy. The moral and epistemic distinction is also used in the literature. Morality refers to the rightness or goodness of actions and epistemic refers to knowledge and truth. Fricker (2007), for example, discusses the moral and epistemic dimensions of injustice.

[3] For theories combining procedural and epistemic elements into democratic legitimacy include Chambers (2017), Estlund (1997) and (2008), Kelly (2012), Landemore (2012) and (2017), MacGilvray (2014), Misak and Talisse, (2014), Mladenović (2020), Muirhead (2016), Peter (2007), (2008), and (2010). I have developed a theory that combines procedural and epistemic dimensions of democratic legitimacy in Min (2014) and (2016). The main arguments in this paper are owed to Min (2014). In deliberative democracy literature, there is an agreement that democratic legitimacy is intimately tied to the epistemic quality of deliberation and its outcomes. Deliberative democrats express the relationship between legitimacy and the epistemic quality of deliberation in various ways. Cohen writes that "outcomes are democratically legitimate if and only if they could be the object of a free and reasoned agreement among equals" (Cohen 1997a: 77). Habermas writes that "The democratic principle states that only those statutes may claim legitimacy that can meet with the assent of all citizens in a discursive process of legislation that in turn has been legally constituted" (Habermas 1996: 110). Mansbridge et al. (2012) writes, "the last several decades have seen growing agreement among political theorists and empirical political scientists that the legitimacy of a democracy depends in part on the quality of deliberation that informs citizens and their representatives" (Mansbridge et. al. 2012). They all express the basic point that the legitimacy of democracy depends, in part, on the quality (rational) of deliberation (Peter 2010). Chambers (2017) expresses this point even more strongly: "deliberative democracy ties legitimacy to the twin values of epistemic quality and equal participation" (Chambers 2017: 9).

[4] There are three main issues in the epistemology and democracy (or epistemic democracy for short) literature. The first issue concerns democratic legitimacy and authority. This is the topic of this paper. The second issue concerns the problem of epistocracy. Epistocracy is a political regime type that authorizes the wise to rule over the many. Various epistocratic proposals have been suggested in the history of political thought, beginning with Plato. More tempered versions can be found in Brennan (2016), Lippman (1925), Mill (1991), Schumpeter (1950). The part of the motivation for epistemic defenses of democracy is to overcome epistocracy. For important responses to epistocracy, see Estlund (2008), Landemore (2012), and Lafont (2020), among others. The third issue concerns the epistemic mechanisms, such as the Condorcet Jury Theorem, the Diversity Trumps Ability Theorem,

ship between the procedural and epistemic dimensions of democracy vis-à-vis democratic legitimacy.[5]

This paper argues that there are two concepts of the epistemic value of public deliberation in deliberative democracy literature, and recognizing this difference will help us to think more carefully about democratic legitimacy. I argue for this thesis in four steps. First, two concepts of the epistemic value of public deliberation exist in the current literature. The strong concept entails a procedure's ability to achieve a correct outcome, as judged by a procedure-independent standard of correctness. The weak concept entails a procedure's ability to facilitate an exchange of reasons. Second, I consider two prominent epistemic theories of democratic legitimacy. David Estlund argues that democratic legitimacy is a function of a procedure's ability to produce correct decisions, as judged by a procedure-independent standard of correctness. Estlund's view combines a strong concept with democratic legitimacy. Fabienne Peter argues that democratic legitimacy is a function of a procedure's intrinsic and constitutive epistemic values. Peter's view combines the weak concept with democratic legitimacy. In my view, both Estlund's and Peter's theories, if combined, provide a sophisticated defense of democratic legitimacy. Combining these two theories, however, requires us to make a distinction between the two dimensions of lawmaking. Thus, I make a distinction between the enactment and long-run dimensions of legitimate lawmaking. Third, the enactment dimension refers to the point when laws are made in the constitutional order. The long-run dimension refers to the evaluation of laws after the enactment of its effectiveness.

Fourth, I combine the two concepts of the epistemic value of public deliberation and two dimensions of legitimate lawmaking in the following ways. The first combination is between the weak concept and the enactment phase of legitimate lawmaking. I argue that a deliberative procedure has weak epistemic values if an exchange of reasons is public knowledge among all citizens. It produces an understanding of the reasons for citizens as political and epistemic agents. The procedure not only generates reasons as political and epistemic agents, but it is also conducive to mutual respect. Thus, the legitimacy of laws at the enactment phase is primarily a function of what survives a deliberative procedure and the resulting laws acceptable to citizens. The second combination is between the strong concept and the long run phase of legitimate lawmaking. I argue that a deliberative procedure has strong epistemic values if a procedure-independent standard of correctness

reliability, et cetera. For this issue, see Anderson (2006), Kelly (2012), List and Goodin (2001), Marti (2006), Talisse (2009), and Goodin and Spiekermann (2018).

[5] This essay does not ask if democracy can be legitimate or what is the best way to legitimize democracy. In other words, I will not engage with philosophical anarchism nor review the well-treaded theories of legitimacy. This essay engages in a narrower question: if democracy can be justified on procedural and epistemic grounds, which theory is the best one to pursue this line of reasoning?

is part of its evaluation of laws. Although laws surviving a deliberative procedure are legitimate at enactment, laws should be evaluated for their long-run consequences. The evaluation of the long-run consequences requires a procedure-independent standard of correctness. Therefore, the legitimacy of laws, in the long run, is primarily a function of satisfying the epistemic criteria of evaluation of consequences.

1. *Strong and weak epistemic value of public deliberation*

In this section, I show that two concepts of the epistemic value of public deliberation exist in the current literature in epistemic and deliberative democracy.[6] The first concept of the epistemic value of public deliberation implies a procedure's ability to achieve a correct outcome, as judged by a procedure-independent standard of correctness. Let us call this the *strong* epistemic value of public deliberation.[7] There is a diversity of sources for this concept, but Joshua Cohen's conception of epistemic democracy is the *locus classicus* of this literature:

> An epistemic interpretation of voting has three main elements: (1) an independent standard of correct decisions – that is, an account of justice or of the common good that is independent of current consensus and the outcomes of votes; (2) a cognitive account of voting – that is, the view that voting expresses beliefs about what the correct policies are according to the independent standard, not personal preferences for policies; and (3) an account of decision-making as a process of the adjustment of beliefs, adjustments that are undertaken in part in light of the evidence about the correct answer

[6] The concept "epistemic" and its cognates refer to truth, justification, reason, reason-giving, knowledge pooling, knowledge production, knowledge transmission, and so forth. Philosophers make fine distinctions between concepts, and I demonstrate that there are two concepts of the epistemic value of public deliberation extant in the literature. I owe the distinction between the strong and weak concept of epistemic to Goldman (1988), but obviously appropriating the idea for my purpose. According to Goldman, the strong justification refers to S believing that p only if p is produced by a reliable belief-forming process. S is weakly justified in believing that p only if S has a good reason to believe p. By "two concepts of the epistemic value of public deliberation," I am distinguishing between 'concept' and 'conceptions' of 'epistemic value of public deliberation.' There are two concepts of the epistemic value of public deliberation, strong and weak, but there might be multiple strong and weak conceptions. The distinction between 'concept' and 'conception' comes from Rawls who writes that it is "natural to think of the concept of justice as distinct from the various conceptions of justice and as being specified by the role which these different sets of principles, these different conceptions, have in common." (Rawls 1971: 5–6). One might object that the distinction between the strong and weak concepts of epistemic value of public deliberation already exists in the literature by way of perfect/imperfect proceduralism on the one hand, and pure proceduralism, on the other hand. The purpose of this paper is to draw that distinction to some logical conclusions.

[7] It is worth noting that most deliberative democratic theorists connect the concept epistemic with such an objective standard of correctness (Estlund 2008; Peter 2008).

> that is provided by the beliefs of others. Thus, the epistemic conception treats processes of decision-making as, potentially, rational processes of the formation of common judgments. (Cohen 1986)

Based on Cohen's description, we can extract three theses for epistemic democracy. The first is political cognitivism, which is a thesis that there are true or false answers or better or worse decisions in politics (Landemore 2012, 2017). The second is that there are true or false answers to politics presupposes a procedure-independent standard of correctness, that is other than rational consensus or consent.

Based on these two theses, the third thesis is that democratic procedure "tracks the truth" better than its alternatives. According to List and Goodin:

> For epistemic democrats, the aim of democracy is to 'track the truth.' For them, democracy is more desirable than alternative forms of decision-making because, and insofar as, it does that. One democratic decision rule is more desirable than another according to that same standard, so far as epistemic democrats are concerned. (List and Goodin 2001: 277)

Fabienne Peter articulates the conceptual connection between an independent standard of correctness and truth-tracking:

> By the standard account I shall denote any characterization of epistemic democracy which centers on the truth-tracking potential of democratic decision-making processes, and in which truth refers to a procedure-independent standard of correctness. According to such accounts, there exists, independently of the actual decision-making process, a correct decision – for example the one that "truly" realizes justice, or the one that is the "true" common good – and the legitimacy of democratic decisions depends, at least in part, on the ability of the decision-making process to generate the correct outcome. (Peter 2008: 33–4)

The most prominent articulation of the truth tracking capacity of deliberation is demonstrated by Landemore (2012), who suggests that deliberation can make correct decisions by harnessing the cognitive diversity of citizens. Aristotle, the earliest exponent of the 'wisdom of the crowd' thesis, argues that deliberation among all is epistemically superior to deliberation among the few.[8]

The second concept of the epistemic value of public deliberation implies that the epistemic value is intrinsic to and constitutive of a procedure as an exchange of reasons. Let us call this the *weak* epistemic value of public deliberation. Again, there is a diversity of sources for this concept, but we can find traces of this concept in at least two places. Joshua Cohen, in his classic article on democratic legitimacy, writes about the common good:

[8] Aristotle (1998). This insight has also been argued by Rawls: "The benefits of discussion lie in the fact that even representative legislators are limited in knowledge and the ability to reason. No one of them knows everything the others know, or can make all the same inferences that they can draw in concert. Discussion is a way of combining knowledge and enlarging the range of arguments" (Rawls 1971: 358–9).

> …[T]he relevant conceptions of the common good are not comprised simply of interests and preferences that are *antecedent* to deliberation. Instead, the interests, aims, and ideals that comprise the common good are those that survive deliberation… (Cohen 1997a: 77. Emphasis mine)

"The ideals that comprise the common good," thus, "are those that survive" a rational deliberative procedure. In other words, the ideals of the common good are constructed through a deliberative process. In a different passage, Cohen argues, "Outcomes are legitimate if and only if they could be the object of a free and reasoned agreement among equals" (Cohen 1997a: 73). Cohen ties the legitimacy of outcomes to properties of the deliberative procedure, such as equality, freedom, and rationality. Cohen's ideal proceduralism recognizes that the overall quality of deliberation matters, but he defines what a good outcome is solely on procedural considerations. Cohen's version of ideal deliberative proceduralism could endorse the second interpretation of the epistemic value of public deliberation because the ideals grounding deliberative values, such as the common good, are constructed within a deliberative procedure and do not refer to a procedure-independent standard. In short, the true common good is constitutive of what survives an ideal deliberation.

More recently, Simone Chambers (2017) argues that that the epistemic ideal of reason-giving has always been with deliberative democracy. Reason giving in deliberative democracy – giving reasons that all could accept – captures the moral requirement of treating citizens as equals. It also captures the epistemic ideals of improving the epistemic quality of decisions. She argues that the "criterion of procedurally-independent standard of correct outcome is not the best way to conceptualize that epistemic dimension of much of deliberative democracy because so much of that dimension is invested in good procedures" (Chambers 2017: 63). Instead, she proposes the feedback loop between the system and the lifeworld, citizens and institutions, and between lay knowledge and social scientific knowledge, as the basis of a good procedure.

In short, I have demonstrated that there are two concepts of the epistemic value of public deliberation operative in the literature. First, epistemic values come from a procedure's ability to achieve a correct outcome, as judged by a procedure-independent standard of correctness. Second, epistemic values are intrinsic and constitutive features of a procedure as an exchange of reasons.

2. *Hybrid views*

Legitimacy is one of the central normative concepts for deliberative democracy because of the fundamental belief that any laws or policies must be mutually justifiable to all. The normative conception of legitimacy – as opposed to the empirical conception that refers to the psychological acceptance of a rule – regards the moral permissibility of a

regime's use of coercive power.[9] Although there is a diversity of views and interpretations of what it means to have democratic legitimacy, let me focus on David Estlund's and Fabienne Peter's views, as their views are the most developed in this arena.[10]

David Estlund, in his groundbreaking book *Democratic Authority,* argues that democracy tends to produce correct or just decisions better than random, and it is better than non-democratic alternatives acceptable from the standpoint of public reason. Estlund's idea is that democracy is epistemically superior to all other political arrangements that are *fair*. Since fairness can be achieved procedurally, Estlund's theory appeals to both procedural and epistemic considerations. He denies that purely procedural considerations are sufficient to judge the legitimacy of the outcome; instead, he argues that procedure-inde-

[9] In the legitimacy literature, there is a tendency to separate legitimacy from authority. Legitimacy means the moral permissibility to enforce a system of rules, whereas authority means the moral right to rule and a corresponding duty to obey. For this essay, I only discuss legitimacy. When it comes to legitimacy, philosophers usually make a distinction between normative and descriptive legitimacy. Descriptive legitimacy is the psychological acceptance of a ruling regime. Normative legitimacy is the moral permissibility to use coercive power. In the domain of normative legitimacy, there are three types of views. Here I follow Habermas (1996) where he articulates the three normative models of democratic legitimacy. The first view is the liberal view, whose progenitors are Hobbes and Locke. For the Lockean view, see Simmons (2008). The second type is the civic republican version of Rousseau. For the republican view, see Pettit (2012). The third type is deliberative theories of legitimacy that can be found in Benhabib (1994); Chambers (2003); Cohen (1997a); Dryzek (2010); Habermas (1996); Lafont (2014) and (2020); Manin (1987); Thompson (2008).

[10] I chose David Estlund's and Fabienne Peter's accounts of democratic legitimacy vis-à-vis the epistemic value of public deliberation because their views, to the best of my knowledge, are the most comprehensive and thoroughly developed in this area. I decided to engage with their earlier works, Estlund (2008) and Peter (2007), (2008), and (2010), because while each philosopher has advanced well past thinking about democratic legitimacy, their earlier works are most relevant to the topic of this paper. Estlund's latest writings focus on the role of ideal theory in political philosophizing. His latest book, *Utopophobia*, is an attempt at such theorizing. See Estlund (2020). Peter, to the best of my knowledge, is still expounding her powerful views on political legitimacy. However, the focus has changed to broader issues in political philosophy and social epistemology. For instance, her (2013) paper discusses peer disagreement and the relevance of the second personal standpoint to illuminate what it means when peers disagree in epistemic and practical deliberations. Her (forthcoming) paper, "Epistemic Norms of Political Deliberation," is primarily about how well-ordered epistemic norms of political deliberation contributes to the political legitimacy of deliberative democracy. It is primarily about epistemic norms, including "procedural epistemic norms" (see section 6 of that paper), and only derivative about how the procedural epistemic norms affect political legitimacy. Her (forthcoming) paper combines the instrumental benefits (vis-à-vis, epistemic benefits) of public deliberation with procedural epistemic norms. This view is philosophically interesting and worth contending with. However, that would be outside the scope of this paper. Moreover, combining the two still requires a distinction between the two concepts of the epistemic value of public deliberation and the two phases of legitimate lawmaking.

pendent epistemic norms must be met, and fair procedures followed if democratic outcomes are to be judged legitimate.

Fabienne Peter, in a series of papers, proposes an alternative epistemic theory of democratic legitimacy that conceptually separates legitimacy from any procedure-independent standard of correctness. Explicitly rejecting such a standard, Peter argues that even if there is no standard to judge the correctness of an outcome, epistemic values can be attributed to the procedure. Peter's pure epistemic proceduralism holds that "legitimate decisions are those which are the result of deliberation under conditions of political and epistemic fairness" (Peter 2008: 50). Thus, the legitimacy of decisions is solely a function of a fair deliberative procedure. There is no other standard for judging the legitimacy of decisions, except the fairness of the procedure. It is defined as one that gives each deliberator an equal chance to express herself, and this opens the door to epistemic considerations.

These two theories represent two distinctive ways to conceptualize the relationship between democratic legitimacy and the epistemic values of public deliberation. Whereas Estlund's theory posits the existence of a procedure-independent standard of correctness, as part of its legitimacy, Peter's view does not posit a procedure-independent standard of correctness. Both theories are necessary components of deliberative democracy literature. The deliberative democratic literature has "come of age" (Bohman 1998) and has now made a "systemic" turn.[11] In the systemic turn, having two concepts of the epistemic value of public deliberation is useful. In some instances, deliberations will be conducive to generating reasons that can enlarge the range of reasons that are useful in a dialogue, without necessarily affecting the deliberative outcomes (Cohen 1997a). In this stage, the weak epistemic concept of public deliberation is operative and useful. In other instances, the goal is to track the truth, meaning whether some deliberative outcomes track empirical facts 'on the ground' or local practical truths. It gives us a 'critical edge' of democratic deliberation over other methods of fact-finding. Such an evaluation requires the strong epistemic concept of public deliberation. Thus, both epistemic concepts of public deliberation are useful to theorize about the epistemic dimension of deliberative democracy.

At the same time, we cannot endorse both concepts without inconsistency. The reason for this is that the strong and weak concepts are potentially inconsistent, for the former posits the procedure-independent standard of correctness and the latter does not. This can be shown by examining three types of theories that combine procedural and epistemic dimensions of democratic legitimacy. First, pure epistemic proceduralism argues that legitimacy is a function of procedural fairness and the epistemic value of deliberation (Peter 2008). Second, perfect epistemic proceduralism asserts that outcomes are infallible, as

[11] For the systemic turn in deliberative democracy, see Mansbridge et. al. (2012) and Owen and Smith (2015) for reviews.

judged by a procedure-independent standard (Rousseau 1984). Third, imperfect epistemic proceduralism argues that the procedure is likely to give correct outcomes though the procedure is fallible, as judged by a procedure-independent standard (Estlund 2008).

According to the perfect and imperfect proceduralist views of democratic legitimacy, the procedure-independent standard of correctness is necessary for the legitimacy of outcomes. On the other hand, according to the pure epistemic proceduralist view of democratic legitimacy, no procedure-independent standard of correctness is necessary to judge the outcome. Hence, Estlund's and Peter's views are both epistemic theories of democratic legitimacy, yet they differ on the role that the epistemic value of public deliberation plays in the process of justification. To put it differently, according to Estlund's view, the epistemic value of public deliberation entails that there is a correct answer, as judged by a procedure-independent standard. According to Peter's view, the epistemic value of public deliberation entails that there is a correct answer, but the correct answer is intrinsic and constitutive of the procedure. In my view, both views are correct, yet they are inconsistent views; that is, we cannot consistently endorse *both* imperfect procedural legitimacy and pure epistemic procedural legitimacy. Is there a way to resolve the tension?

The first way to resolve the tension is to think of legitimacy primarily as the legitimacy of the procedure (due to the procedure's tendency to get it correct), which then *confers* legitimacy on particular outcomes. Hence, the outcome legitimacy is a *derivative* of procedural legitimacy.[12] This appears to be Estlund's way of thinking.[13] The basic idea here is that the legitimacy of democracy is a function of the procedure's tendency to arrive at correct decisions and fair procedures. What confers legitimacy on individual outcomes of the jury trial is not the correctness of individual outcomes themselves, but the epistemic values of the jury trial – evidence gathering, testimony, fair trial, et cetera. These procedural elements are what confers legitimacy on individual decisions. Analogously, what confers legitimacy on individual outcomes in a democracy is not the quality of individual outcomes, but the democratic procedure – voting, deliberation among diverse perspectives, and the like. Essentially, this is a two-level view: the procedure on the structural level enjoys imperfect procedural legitimacy, whereas outcomes enjoy purely procedural legitimacy.

The second way to resolve the tension between imperfect procedural legitimacy and pure procedural outcome legitimacy is "pure epistemic proceduralism." Pure epistemic proceduralism argues that "legitimate decisions are those which are the result of deliberation under condi-

[12] See Rehg (1997), for an explication of this view.

[13] Estlund writes: "[Epistemic proceduralism] is a proceduralist view, linking legitimacy and authority of a decision to its procedural source and not to its substantive correctness" (Estlund 2008: 116).

tions of political and epistemic fairness" but "legitimacy does not depend on procedure-independent standards of correctness, or on their ability to contribute to the common good by solving social problems" (Peter 2008: 50). In other words, legitimate laws are solely a function of purely procedural considerations of political and epistemic fairness (Peter 2008, 2010).

My proposal then is to make a distinction between the two dimensions of legitimate lawmaking that can accommodate both concepts of the epistemic value of public deliberation. This is proposed as an ecumenical move that can appeal to both views.

3. *Two-dimensions of legitimate lawmaking*

This section explains that there are two dimensions of legitimate lawmaking. The first dimension, the enactment dimension, is the process of laws' enactment in a democracy. In this dimension, reasons for legal proposals are debated and traded. In formulating public opinions in the informal public sphere, every relevant viewpoint should be included in the discourse. Given that the norm of deliberation is the "forceless force of the better argument,"[14] it is the job of the participants to sift reasons and weigh the evidence. It is also incumbent upon deliberators to filter reasons and viewpoints that are irrelevant to the discourse at hand. This process also includes the process of filtering of acceptable viewpoints entering the formal institutions of the state.

The second dimension is the process of evaluating the long-term consequences of laws. Our evaluation of injustice and oppression is sometimes attributed to the negative consequences. While the consequences themselves are not strictly relevant to the legitimacy of laws, how the polity assesses and modifies laws and policies have relevance to the legitimacy of laws.

Three points must be kept in mind. First, the distinction between enactment and long-run dimensions is *analytical*. The distinction may not always be cut and dry in the real-life decision-making process. Second, the distinction between enactment and long-run dimensions is a temporal notion. Take a simple example: suppose Tom wants to lose weight and adopts a workout regimen. After six months of working out, Tom lost 20 pounds. Whether the workout regimen works can be judged in retrospect. Just in the same way, lawmaking has a temporal dimension. Before the law is enacted at time *t1*, the law goes through whatever deliberative and democratic process it needs to go through.

[14] This phrase is originated by Jürgen Habermas in various places, but here is a direct passage from his (1999) paper. He writes that the discourse, "which is supposed to be public and inclusive, to grant equal communication rights for participants, to require sincerity and to diffuse any kind of force other than the forceless force of the better argument. This communicative structure is expected to create a deliberative space for the mobilization of the best available contributions for the most relevant topics" (Habermas 1999: 3).

Later, at *t2* the effects of laws on citizen attitude, reduction of harm, or the promotion of well-being have become evident. If citizens see at time *t2* that the law enacted at *t1* reduces harm, then we would judge that law to be working well; on the contrary, if citizens see at time *t2* that the law enacted at *t1* increases harm, then we would judge that the law works poorly.

Third, there is a dynamic relationship between the enactment and long-run phases of lawmaking. That is, the lawmaking process is not merely one-directional – the lawmaking progression goes from enactment to the long run – but there is a dynamic relationship between the two. According to Anderson,

> [V]oting and deliberation represent alternating moments in a continuous process of provisional decision-making, the aim of which is simultaneously to learn about what works and to decide upon criteria of what counts as working from the perspective of citizens acting and thinking collectively. Decisions are provisional and continuously subject to revision in light of feedback from citizens about their consequences. (Anderson 2009: 217)

Citizens give justifications for endorsing a policy by providing publicly acceptable reasons. But a democratic society also does not consider the matter settled even after we provide justifications. That is because circumstances change, cultures change, and social values and norms change.[15] When they do, democratic citizens should provide reasons for the change, and debate whether the change is desirable. Similarly, intelligent policy and lawmaking require that we anticipate the likely consequences of collective action. Continually revising the means-end relationship in policymaking is one way to accomplish that goal.

If the distinction between two concepts of epistemic value of public deliberation and two phases of legitimate decision-making holds, then we should re-conceptualize the relation between epistemic values of democracy and democratic legitimacy. I begin with the strategy of combining the weak concept of the epistemic value of public deliberation and legitimacy at enactment in the next section, and the following section combines the strong concept of the epistemic value of public deliberation and legitimacy in the long run. Let me start with the first strategy.

4. *The weak concept of the epistemic value of public deliberation and legitimacy at the enactment*

This section argues that the legitimacy of laws at enactment is primarily a function of what survives a deliberative procedure. I offer four reasons to support this conclusion. First, a deliberative procedure has weak epistemic values if an exchange of reasons is public knowledge among all citizens. A deliberative procedure is fair when citizens' in-

[15] While moral norms are valid from a universal point of view and transcend historical time and social space, I think political and ethical norms are historical and contextual.

terests, voices, and perspectives are considered in the opinion-and-will-formation processes in relevant ways. In addition to procedural fairness, inclusion is also a constitutive feature of the deliberative procedure. Inclusion typically refers to the all-affected principle, where all those affected by collective decisions must have the opportunity to provide input (Goodin 2007). The all-affected principle emphasizes the dynamic aspects of the constitution of the public, and thus "serves to take normative command of a situation of plural and competing allegiances" (Näsström 2011: 123; Young 2000; Dryzek 2010).

Besides the normative dimension of inclusion, a deliberative procedure also has weak epistemic values. Every citizen occupies some social space. Occupying social space from an isolated corner of existence garners only a narrow understanding of the complexity of how the world is constituted. Citizens, because of their life histories and stories, have much to offer in filling in the content of perspectives. This perspectival knowledge comes from occupying social space and functioning, occupation, gender roles, religious affiliation, and so forth. By learning from differently situated others, citizens collectively come to an enlarged understanding of the social world. It means that social knowledge is not something that one individual or social group can achieve by themselves. Social knowledge is possible only through the pooling of diversely situated knowledge. Hence, social knowledge is achievable only within the context of an inclusive deliberative process among diverse perspectives. It is through the process of justifying one's reasons to others that creates a diverse pool of perspectives. Since citizens come from different backgrounds and life situations, such a deliberative process will result in having a diverse pool of perspectives. The pooling of situated perspectives then allows us to have a mutual understanding of *ideas about* the world that everyone dwells in together.[16] Notice here that the social knowledge gained through deliberation is not a set of fixed points that can be gotten through an impartial deliberation. Rather, social knowledge is *constructed* through deliberation, meaning that social knowledge is not something that exists before deliberation; social knowledge is discovered or constructed through deliberation. Moreover, ends are not fixed points, but they are revisable through deliberation. Social knowledge and objectivity are, therefore, the product of system-wide deliberation among free and equal citizens. Thus, a deliberative procedure has weak epistemic values if the process of reason-giving is a public knowledge among all citizens.

Second, the epistemic value of public deliberation that flows out of the deliberative procedure generates a public understanding of citizens' reasons and interests by political and epistemic agents. Citizens are not only political agents, insofar as they have certain specified rights and duties associated with citizenship, but they are also epistemic agents. That is, citizens usually possess politically relevant practical knowl-

[16] I draw from Young (2000), Chapters 3 and 4.

edge. Firstly, citizens possess knowledge about their "enlightened interests" – their self-interests and how to execute them.[17] Secondly, citizens have to know they can make relevant contributions to a discourse. Sometimes citizens do not know that they can contribute. Many citizens are oppressed, due to poverty, institutional racism, and marginalization. Because of that long-standing oppression, they do not believe that they have anything to offer. Hence, knowing that one is able and entitled to contribute to discourse is crucial. Moreover, citizens have to know that their contribution can be relevant to a discourse. This takes some skills as a citizen because it is not immediately obvious when one has a relevant contribution.[18] Thirdly, they have factual knowledge arising from one's social roles. For example, an employee at Wal-Mart making the minimum wage knows that he cannot survive on that wage even if he works 40 hours. This information seems politically relevant because it will be helpful to know whether the minimum wage is a sustainable wage for workers in a location.

Given these three kinds of politically relevant knowledge, the political and epistemic inclusion as a procedure is a way in which we respect citizens as political and epistemic agents. This expresses the value of the procedural principle of political and epistemic inclusion: we ought to appreciate that citizens are political and epistemic agents and they ought to be respected as such. In other words, citizens should be respected as political and epistemic agents and not merely for the instrumental epistemic benefits of arriving at correct outcomes.

Third, a deliberative procedure that has the weak epistemic value of public deliberation not only generates a public understanding of reasons as political and epistemic agents, but it also generates mutual respect. Having one's voice heard is a way of being respected. On the individual agency level, one feels respected when someone listens to their concerns, and their reasons are considered and listened to. Any mature person knows that one cannot get one's way all the time or even most of the time. However, having one's voice heard with sincerity by others is enough sometimes. Conversely, not having my voice heard is a sign of disrespect and non-recognition.

Fourth, laws produced by a deliberative procedure are likely to be reasonably acceptable to all citizens at the moment of enactment. Because reasons have been discussed and traded in the deliberative process, one can reasonably accept the outcome if one's reasons are discussed and not lightly dismissed. That is, human beings have a right to justification. The right to justification entails that it is procedurally required to have a right to have a say and listened to, but it does not mean that they will be taken into consideration. Thus, our right to

[17] See Mill (1991) for a discussion about the enlightened interests of the citizens.

[18] The connection between this sentence to epistemic injustice is salient, though it is outside the scope of this paper. See Fricker (2007) and Fricker (2015) for discussion about epistemic injustice.

justification is sufficient, sometimes, for our acceptance of the outcome.

Premises three and four are crucial steps in my argument. Fairness and inclusion are not only procedural values, but they are also epistemic values. The epistemic values are constitutive of the procedure and not independent of them. Stated differently, we should not divorce the procedural values of fairness and inclusion from their epistemic features. It means that a procedure has constitutive epistemic features, but we should conceptually separate the procedure-independent standard of correctness from the meaning of epistemic. Thus, laws are likely to be acceptable to all citizens because everyone's opinions, perspectives, and interests were considered. This last point will be controversial, so let me explain.

The reason why the acceptability of laws is dependent on the inclusion of everyone's perspectives is that our normative attitude against laws is a function of mutual respect. The principle of mutual respect explains the normative attitudes against the acceptability of laws. One's normative attitude against laws is that if one's voice was respectfully considered, then one will likely accept the law's legitimacy. Even if one disagrees, the public nature of the deliberative procedure will inform that one's reasons have been considered. Sometimes, it is the prospect of being listened to and taken seriously rather than the outcome themselves. Thus, what matters primarily is the procedure, not the substantive or epistemic features.

This normative point about mutual respect can be linked up with empirical findings in the social psychology of procedural justice (Lind and Tyler 1988). The basic idea behind the research is this: it is widely presumed that "outcome justice" – people's attitude towards the justice of outcomes – is sufficient to motivate people to comply with fair outcomes. But the motivation behind procedural justice is that fair process is important, and in some cases, it is more important than the outcome. The process is important because, in a good process, people feel that they have been treated fairly; they feel like the process is fairer; that, in turn, *could* increase their satisfaction with the outcome, whether favorable or not (Delli Carpini et. al. 2004: 327). Allan Lind and Tom Tyler describe the intuition behind this thinking by recounting a story:

> Judges in that [traffic court in the city of Chicago] often take the view that showing up for court and losing a day's pay at work is punishment enough or a traffic offense. As a result, those who arrive in court often have their case dismissed without any hearing. From a defendant's perspective this is a good outcome—the defendant pays no fine, does not go to jail, and has no violation record. However, interviews with traffic court defendants suggest that despite these favorable outcomes they often leave the court dissatisfied. For example, one woman showed up for court with photographs that she felt showed that a sign warning her not to make an illegal turn was not clearly visible. After her case was dismissed (a victory!) she was angry and expressed considerable dissatisfaction with the court (as well as making several unflattering remarks about the judge). (Lind and Tyler 1988: 2)

Lind and Tyler go on to say that the outcome-based models "might find the woman's dissatisfaction difficult to explain, but process-based models would have little trouble in accounting for her reaction: the woman felt angry because the outcome she received was not arrived at using a procedure that met her standards of proper judicial process" (Lind and Tyler 1988: 2). This kind of judgment and intuition has been mostly studied in law, but the authors argue that it is generalizable to other areas of social interactions including politics and lawmaking (Lind and Tyler 1988: chapters 6 and 7).

This empirical evidence demonstrates that the woman cared more about how she was treated by the judge than the favorable outcome. Although Lind and Tyler do not quite put it this way, we can reasonably say that the woman felt disrespected for not being able to present her case to the judge. Sometimes, resolving one's grievances in the right sort of way is equally important as having a favorable outcome.

One might question the validity of the analogical argument between procedural justice and the women's grievance and the weak epistemic value of public deliberation and legitimacy at enactment. It is a bad analogy, one might argue, because the woman felt grieved because the procedure lacked some procedure-independent epistemic qualities such as judicial review.[19] In response, the source of the woman's anger and dissatisfaction arose from the fact that the procedure lacked some procedural qualities, not because the procedure was missing some procedure-independent epistemic quality. She felt angry, and the source of her dissatisfaction with the court was that she was not able to express herself to the judge, even if the decision was made in her favor. In other words, the reason why she was angry and dissatisfied was that she was not heard, despite the favorable outcome. Being able to express one's grief and grievances is one of the rights provided to citizens, at least in the United States. Thus, several notable institutions make hearings, both formal and informal, possible on many levels of government. For instance, one can appeal their social security benefits with the United States government through a congressional or senate field office. One can partake in public hearings on government projects.

This section argued that the legitimacy of laws at enactment is primarily a function of what survives a deliberative procedure. A deliberative procedure has weak epistemic values if an exchange of reasons is public knowledge among all citizens. It generates an understanding of the reasons for citizens as political and epistemic agents. An inclusive procedure that is weakly epistemic not only garners an understanding of reasons as political and epistemic agents, but it also generates mutual respect. Thus, laws produced by the deliberative procedure are legitimate at enactment when binding enforceability is reasonably acceptable to all citizens.

[19] See Goldman (2001) about the role of expertise in democratic societies.

5. *The strong concept of the epistemic value of public deliberation and long-run legitimacy*

The previous section has argued that the legitimacy of laws at enactment is primarily a function of what survives a robust deliberative procedure. In this section, I argue that the legitimacy of laws, in the long run, is a function of their consequences as evaluated by a procedure-independent standard of correctness. I offer four reasons for this conclusion.

Firstly, a deliberative procedure has strong epistemic values if a procedure-independent standard of correctness is part of its evaluation of laws. Recall that the strong epistemic value of public deliberation implies that we cannot separate the epistemic values from a procedure-independent standard.

Secondly, although laws surviving a deliberative procedure are legitimate at enactment, laws must be evaluated for their long-run consequences. There are two reasons why the evaluation of the long-term success of decisions should be part of their legitimacy. The first reason is that we will not know which laws are best or better until they hold up in practice in the long run. Because of the fallibility of human cognition and human institutions, and historical lessons of the past, we do not know what the effects of a legal-political outcome would be until its enactment. Evaluation of the long-term success of laws requires the knowledge of the consequences of enacting a policy or law (both intended and unintended but foreseeable), attitudes of citizens toward the law, as well as effects on institutions, practices, social conditions, and social dynamics. The importance of this knowledge is that laws and policies affect citizens in consequential ways. Laws have effects (good, bad, or neutral); laws can benefit or harm its citizens living under them. Even if one accepts my argument that the long-run success of laws matter, the controversial step in my argument will be whether the legitimacy of laws has anything to do with their long-run success.

This leads us to a second consideration that political justification is provisional. Amy Gutmann explains the provisional nature of political justification:

> [T]he legitimate exercise of political authority requires justification to those people who are bound by it, and decision-making by deliberation among free and equal citizens is the most defensible justification anyone has to offer for provisionally settling controversial issues…[t]he first advantage of deliberative democracy [against competing defenses of democracy] is its recognition of the provisional nature of justification in politics. The empirical and moral understandings of citizens change not only over time and social space but also by virtue of deliberative interchange, the give-and-take of sometimes complementary, often conflicting, political insights and arguments. (Gutmann 1996: 344)

Deliberative democracy holds that one of the central criteria of legitimacy is mutual justification among free and equal citizens (Chambers

2003; Lafont 2014; Thompson 2008). The coercive nature of political authority has to be justified vertically from the state to citizens and horizontally between citizens by giving reasons for why one endorses such a policy. Public reasons, or publicly justifiable reasons, are informed by cultural differences and changes, changes, in fact, social values, and changes in social norms. When those changes occur, as they inevitably do over time, democratic citizens should debate whether the change is desirable.

The theoretical basis of this is the ideals of dissent. Dissent is a contestatory mechanism to protect minority rights against the possibility of the tyranny of the majority. It also enables citizens to combat unjust and unintelligent laws. In liberal democratic societies like the United States, for example, there are periodic elections on both federal and local levels to ensure the possibility of repealing laws. While those feedback mechanisms of liberal democracy are institutional safeguards against the majority's tyranny, such liberal mechanisms (guaranteed through the constitution) are not enough. Deliberative democrats usually go one step further to emphasize the constitutive importance of contestation. Because deliberation is essentially open-ended, any policy option or choice can be contested from a variety of perspectives. In short, a democratic society does not consider the matter settled once justifications are given.

Thirdly, the evaluation of the long-run consequences requires a procedure-independent standard. There are two reasons for this. First, politics entails deep moral disagreements. These moral disagreements are not merely conflicts of power, desire, or preference. They are sometimes disagreements in worldviews, philosophies, and ideologies. Adjudicating them requires some standard as to which choices are more or less superior. That evaluation makes little sense if we do not posit that there are some standards by which we evaluate solutions based on some procedure-independent standard of correctness. Suppose public deliberation does not aim at truth from some procedure-independent standard of correctness. When there is a political disagreement we would not even be able to inquire whether X's position or Y's position is correct. If that evaluation cannot be had, then the disagreement becomes X's expressing X's desires and preferences and Y's expressing Y's desires and preferences. Thus, deliberation reduces down to a mere power struggle. In cases of minor disputes – whether to give to the Catholic Charities or the Oxfam or to serve strawberry or chocolate ice cream at the local fundraising—X's desire and Y's desire can be subject to reasonable disagreement. In cases of major disputes, however, we should be able to say X is correct, and Y is incorrect. For example, in our deliberation about whether the drone attacks are a justifiable form of killing, we should be able in principle to judge that X's position is correct, and Y's position is incorrect. If it were not possible to say X is correct, and Y is incorrect, then there would be little reason to

deliberate. Therefore, public deliberation must aim at truth from some procedure-independent standard of correctness.

Second, a procedure-independent standard is necessary to judge whether the outcome is good or not. That evaluation requires some procedure-independent standard of correctness. Some laws are morally weightier than others for sure, but it seems undeniable that all laws are subject to evaluation. Law's efficacy, which is an evaluation of the law's success or failure, depends on some objective standard, in the sense that there are measurable consequences independently of an actual procedure or a current consensus. In other words, the amelioration of the present condition requires an objective standard for evaluation.

The provisional nature of political justification provides a normative reason why collective outcome should be subject to continual and ongoing justification, especially when the law fails to serve the shared interests of all citizens. Although laws and policies surviving a robust deliberative procedure are legitimate laws at enactment, we will not know which laws are best until they hold up in practice in the long run. The evaluation of which have long term success requires a retrospective judgment. Making a retrospective judgment about the long-term success of a legal-political outcome involves the strong epistemic values of democracy – democracy tending to produce substantially correct decisions judged from a procedure-independent standard of correctness. If my argument is on track, then we should consider the long-run success of laws in legitimizing collective outcomes.

This section has argued that the legitimacy of laws, in the long run, is primarily a function of satisfying the epistemic criteria of evaluation of consequences. To support this conclusion, I have argued that a deliberative procedure has strong epistemic values if a procedure-independent standard of correctness is part of its evaluation of laws. Although laws surviving a deliberative procedure are legitimate at enactment, we must evaluate laws' long-run consequences. The evaluation of the long-run consequences requires a procedure-independent standard.

6. *Answering objections*

Before concluding, I wish to address four objections. First, one might object that the two-dimensional analysis of democratic legitimacy is a form of democratic instrumentalism or consequentialism about democratic legitimacy. To say that the legitimacy of outcomes must be evaluated against some standard is to invoke democratic instrumentalism. My response is that the second phase of legitimate law-making is closely related to democratic instrumentalism because it argues that citizen acceptance of laws and policies are dependent on their long-run effectiveness. But my view is not *merely* democratic instrumentalism because the first dimension of legitimate lawmaking is a version of pure proceduralism about legitimacy. After all, there is no need to appeal to legal or political consequences in citizen acceptance of policies.

Second, some might object that some laws could be legitimate or illegitimate *a priori*. That is, some laws legitimate or illegitimate, at the time of their enactment, regardless of their consequences. Some laws are illegitimate because they violate individual liberty or freedom. For instance, one might argue that laws that permit discrimination based on one's race or gender are illegitimate regardless of knowing their consequences. If this objection works, then it would undermine the thesis that some deliberative outcomes change its legitimacy status as time goes on.

While I grant that some laws could be legitimate or illegitimate *a priori*, we can question the legitimacy of some laws once the long-run consequences of laws become known. Let me offer a real-life example to articulate this point.

The 18th amendment prohibited alcohol consumption and distribution.[20] There are various political and social impetuses behind Prohibition. On the political front, thirteen states by 1850 had enacted alcohol prohibition legislation. Also prominent in this period was the single-issue lobbying of the Anti-Saloon League (ASL). ASL's powerful presence at the national level brought about a social and cultural shift in the perception of alcohol. Once the "[p]rohibition became the law of the land, many citizens decided to obey it. Referendum results from the immediate post-Volstead showed widespread support" (Blocker 2006: 237). However, once the negative consequences of Prohibition became widely known, the law began to lose its popular support, eventually leading to its appeal in 1933. The Daily Mirror, on the cover page of December 5, 1933, showed men holding beers to celebrate: "PROHIBITION ENDS AT LAST." Although the 18th amendment that made alcohol prohibition lawful was a legitimate law at enactment, as the time progressed, the law began to lose its popular support once the consequences of the law began to be known. This example demonstrates the thesis of this paper. The legitimacy of laws at enactment is primarily a function of what survives a robust deliberative procedure, whereas the legitimacy of laws, in the long run, depends on their epistemic correctness.

Third, one might argue that the above argument rests on a controversial assumption that the law's qualities will be so clear that all citizens will be able to perceive it as correct or incorrect. This assumption is necessary for the argument to work, yet controversial and problematic. There are two ways to understand the objection. First, there is a distinction between the law's correctness and citizens' perception of the law's correctness. So, it is possible that citizens can be mistaken about the correctness of the laws, or refuse to accept the correctness because of their ideological delusion or something else. Second, there

[20] The National Prohibition Act of 1919, the 18th Amendment to the U.S. Constitution, states that "manufacture, sale, or transportation of intoxicating liquors within, the importation thereof into, or the exportation thereof from the United States and all territory subject to the jurisdiction thereof for beverage purposes is hereby prohibited (U.S. Constitution).

can be reasonable disagreement about the correctness of laws and policies. Rawls (1993) famously argued that the modern world is characterized by reasonable pluralism and that there are bound to be reasonable disagreements on comprehensive doctrines. The best we can do in a modern polity is to seek overlapping consensus on constitutional essentials and matters of justice. Here, Estlund's view is different from Rawls's in one respect: while Rawls believes that there will be reasonable disagreements on most things, Estlund believes that there will be no reasonable disagreements (qualified points of view) about some things. This is what he calls the "primary bads." Genocide and slavery are not subject to reasonable disagreement, and all qualified points of view would agree that they are bad. He argues that democratic procedures can avoid these things better than other fair and inclusive procedures.

I agree with Estlund that some laws and policies should be ruled as bad and there are procedure-independent epistemic qualities about them. However, for most laws and policies, what counts as reasonable disagreement, if there are reasonable disagreements, cannot be ruled out *ex-ante*. Thus, we should follow the deliberative process at least on most issues to clarify what the reasonable disagreement entails. Thus, at the moment of enactment of laws, citizens could reasonably agree that the law had followed procedural steps, but the reasonable disagreement comes *ex post facto*. To further articulate this point, consider the "War on Drugs" in the United States. By most accounts, the "War on Drugs" in the United States is an abject failure (Wacquant 2012). The law does not reduce the supply of drugs nor prevent people from using drugs. The enforcement of drug law is costly; the federal government spends billions of dollars. There is an overflow of non-violent offenders in federal prisons; human capital is sacrificed because people in prison cannot vote, work, nor contribute to society; and social capital is decimated in some subsections of the population. According to the Sociologist Loic Wacquant, African-American males are most profoundly affected by this law for three reasons: they are excluded from the education system, including higher education (cultural capital), from jobs and social mobility (social redistribution), and the voting booth (political participation) (Wacquant 2012: 57–8). The lesson is that the abject failure of one law has caused (either directly or indirectly) tremendous suffering for those affected by it.

Now, what reasons do African-Americans (or anyone else) have for accepting the legitimacy of the drug law? The drug law may have been legitimately made at enactment – it went through proper constitutional, legislative, and deliberative channels. If so, then the binding enforceability would be reasonably acceptable to all. But people still deem the law a failure. The reason why people complain about the law is not its enactment illegitimacy, but the law had ill-effects and does not accord with our sense of justice. The unjust treatment of and the exclusion of a large segment of the population fails to track the in-

terests and ideas of those suffering the consequences. In short, if "the legitimate exercise of political authority requires justification to those people who are bound by it," then the drug law is no longer one of the instances of the legitimate exercise of political authority. The drug law cannot be justified to those people who are bound by it, and people may reasonably reject the law as illegitimate.[21]

Fourth, the above example involving the drug law, one might object, brings out an ambiguity between normative and descriptive uses of legitimacy in this paper. Normative legitimacy means the moral permissibility to enforce a system of rules through coercion, whereas descriptive legitimacy is the psychological acceptance of a ruling regime. Philosophers make a sharp distinction between normative and descriptive legitimacy in their discussions. While I accept the distinction, the normative status of the legitimacy of a law is affected by the long-term consequences of law and citizens' reasonable disagreements and psychological attitudes and acceptance towards the law. If the outcome were shown to be bad over time, then citizens would have reasons to reject it as illegitimate. These two factors are not sufficient to overturn the normative status of legitimacy. Nonetheless, those two factors give us reasons for thinking that the long-run negative consequences and people's attitudes and confidence and trust of a system of rule erode and weakens.

7. *Conclusion*

This paper made the relevant distinctions between the strong and weak concepts of the epistemic value of public deliberation and linked those concepts to two dimensions of legitimate lawmaking. If my ecumenical approach to democratic legitimacy is on the right track, then both pure epistemic proceduralists and epistemic proceduralists can endorse my theory. Pure epistemic proceduralists can accept that a deliberative procedure is weakly epistemic and is most applicable in the enactment phase of the legitimate lawmaking process. Epistemic proceduralists, on the other hand, can adopt the strong epistemic values of public deliberation, which serves as a benchmark of knowing when and how improvements are made.[22]

[21] This is from a normative standpoint. It is a different story about whether people comply with the law. People may comply because of the fear of punishment or being ostracized. Furthermore, I am not arguing that all unjust laws are immediate grounds for non-compliance. But I think some laws are so unjust that it is reasonable for the citizens not to comply.

[22] I would like to thank the anonymous reviewers for their helpful comments on the previous version of this paper.

References

Anderson, E. 2006. "The Epistemology of Democracy." *Episteme* 3: 9–23.

Anderson, E. 2009. "Democracy: Instrumental vs. Non-instrumental values." T. Christiano and J. Christman (eds.). *Contemporary Debates in Political Philosophy*. London: Wiley-Blackwell: 213-229.

Behabib, S. 1994. "Deliberative Rationality and Models of Democratic Legitimacy." *Constellations* 1 (1): 26–52.

Blocker, J. 2006. "Did Prohibition really work? Alcohol prohibition as a public health innovation." *American Journal of Public Health* 96 (2): 233–243.

Bohman, J. 1998. "The Coming of Age of Deliberative Democracy." *Journal of Political Philosophy* 4: 418–443.

Bohman, J. and Rehg, W. (eds). 1997. *Deliberative Democracy: Essays on Reason and Politics*. Cambridge: MIT Press.

Brennan, J. 2016. *Against Democracy*. Princeton: Princeton University Press.

Chambers, S. 2003. "Deliberative Democratic Theory." *Annual Review of Political Science* 6: 307–26.

Chambers, S. 2017. "The Epistemic Ideal of Reason-Giving in Deliberative Democracy." *Social Epistemology Review and Reply Collective* 6 (10): 59–64.

Cohen, J. 1986. "An Epistemic Conception of Democracy." *Ethics* 97 (1): 26–38.

Cohen, J. 1997a. "Deliberation and democratic legitimacy." In Bohman and Rehg (eds.).

Cohen, J. 1997b. "Process and Substance in Deliberative Democracy." In Bohman and Rehg (eds.).

Delli Carpini, M. X, Cook, F. L. and Jacobs, L. R. 2004. "Public deliberation, discursive participation, and citizen engagement: A Review of the Empirical Literature." *Annual Review of Political Science* 7: 315–44.

Dryzek, J. 2010. *Foundations and Frontiers of Deliberative Governance.* Oxford: Oxford University Press.

Estlund, D. 1991. "Who's Afraid of Deliberative Democracy. On the Strategic/Deliberative Dichotomy in Recent Constitutional Jurisprudence." *Texas Law Review* 1437.

Estlund, D. 1993. "Making truth safe for democracy." In D. Copp, J. Hampton and J. Roemer (eds.). *The Idea of Democracy*. Cambridge: Cambridge University Press: 71–100.

Estlund, D. 1997. "Beyond Fairness and Deliberation: The Epistemic Dimension of Democratic Authority." In Bohman and Rehg (eds.): 174–204.

Estlund, D. 2008. *Democratic Authority: A Philosophical Framework*. Princeton: Princeton University Press.

Estlund, D. 2020. *Utopophobia*. Princeton: Princeton University Press.

Fricker, M. 2007. *Epistemic Injustice*. Oxford: Oxford University Press.

Fricker, M. 2015. "Epistemic Contribution as a Central Human Capability." In G. Hull (ed). *The Equal Society*. Lanham: Lexington Books: 73–90.

Goldman, A. 2001. "Experts: Which Ones Should You Trust?" *Philosophy and Phenomenological Research* 63 (1): 85–110.

Goldman, A. 1988. "Strong and Weak Justification." *Philosophical Perspectives* 2: 51–69.

Goodin, R. and Spiekermann, K. 2018. *An Epistemic Theory of Democracy*. Oxford: Oxford University Press.

Gutmann, A. 1996. "Democracy, Philosophy, and Justification." In S. Benhabib (ed.). *Democracy and Difference: Contesting the Boundaries of the Political*. Princeton: Princeton University Press: 340–347.

Gutmann, A. and Thompson, D. 1996. *Democracy and Disagreement*. Princeton: Princeton University Press.

Gutmann, A. and Thompson, D. 2004. *Why Not Deliberative Democracy?* Princeton: Princeton University Press.

Habermas, J. 1996. "Three Normative Models of Democracy." In S. Benhabib (ed.). *Democracy and Difference: Contesting the Boundaries of the Political*. Princeton: Princeton University Press: 21–31.

Habermas, J. 1999. "Between Facts and Norms: An Author's Reflections." *Denver University Law Review* 76 (4): 937–942.

Habermas, J. 2006. "Political Communication in Media Society: Does Democracy Still Enjoy an Epistemic Dimension? The Impact of Normative Theory on Empirical Research." *Communication Theory* 16 (4): 411–426.

Kelly, J. T. 2012. *Framing Democracy: A Behavioral Approach to Democracy*. Princeton: Princeton University Press.

Lafont, C. 2014. "Deliberation, participation, and democratic legitimacy: Should deliberative mini-publics shape public policy." *The Journal of Political Philosophy* 1–24.

Lafont, C. 2020. *Democracy without Shortcuts: A Participatory Conception of Deliberative Democracy*. Oxford: Oxford University Press.

Landemore, H. 2012. *Democratic Reason*. Princeton: Princeton University Press.

Landemore, H. 2017. "Beyond the Fact of Disagreement? The Epistemic Turn in Deliberative Democracy." *Social Epistemology* 31 (3): 277–295.

Lind, A. and Tyler, T. 1988. *The Social Psychology of Procedural Justice*. New York: Plenum Press.

Lippmann, W. 1925. *The Phantom Public*. The MacMillan Company.

List, C. and Goodin, R. 2001. "Epistemic Democracy: Generalizing the Condorcet Jury Theorem." *Journal of Political Philosophy* 3: 277–306.

MacGilvray, E. 2014. "Democratic doubts: pragmatism and the epistemic defense of democracy." *The Journal of Political Philosophy* 22 (1): 105–123.

Manin, B. 1987. "On Legitimacy and Political Deliberation." *Political Theory* 15: 338–368.

Mansbridge, J. and Parkinson, J. (eds.). 2012. *Deliberative Systems: Deliberative Democracy on Large Scale*. Cambridge: Cambridge University Press.

Mansbridge, J., et. al. (2012). "Systemic approach to deliberative democracy." In Mansbridge and Parkinson (eds.): 1–28.

Marti, J. L. 2006. "The Epistemic Conception of Deliberative Democracy Defended: Reasons, Rightness, and Equal Political Autonomy." In Marti and Besson (eds.).

Marti, J. L. and Besson, S. (eds.). 2006. *Deliberative Democracy and Its Discontent*. Burlington: Ashgate.

Mill, J. S. 1991. *Considerations on Representative Government*. *On Liberty and Other Essays*. Oxford: Oxford University Press: 205–470.

Min, J. B. 2014. *An Epistemological Defense of Deliberative Democracy*. Doctoral Dissertation.

Min, J. B. 2016. "Politics Must Get It Right Sometimes: Reply to Muirhead." *Critical Review* 28 (3–4): 404–411.

Misak, C. and Talisse, R. 2014. "Debate: pragmatist epistemology and democratic theory: a reply to Eric MacGilvray." *The Journal of Political Philosophy* 22 (3): 366–376.

Mladenović, I. 2020. "Democracy, Truth, and Epistemic Proceduralism." *Croatian Journal of Philosophy* 20 (59): 71–82.

Näsström, S. 2011. "The Challenge of the All-affected Principle." *Political Studies* 59 (1): 116–134.

Owen, D. and Smith, G. 2015. "Survey Article: Deliberation, Democracy, and the Systemic Turn." *The Journal of Political Philosophy* 23 (2): 213–234.

Peter, F. 2007. "Democratic Legitimacy and Proceduralist Social Epistemology." *Politics, Philosophy, & Economics* 6: 329–353.

Peter, F. 2008. "Pure Epistemic Proceduralism." *Episteme* 5 (1): 33–55.

Peter, F. 2010. *Democratic Legitimacy*. London: Routledge.

Peter, F. 2013. "The procedural epistemic value of deliberation." *Synthese* 190: 1253–1266.

Peter, F. 2014. "Political Legitimacy." *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.), URL=<http://plato.stanford.edu/archives/spr2014/entries/legitimacy/>. Web. 15 O c - tober.2013

Peter, F. Forthcoming. "Epistemic norms of political deliberation." In M. Hannon and J. de Ridder (eds.). *Routledge Handbook of Political Epistemology*. New York: Routledge.

Pettit, P. 2012. *On People's Terms*. Oxford: Oxford University Press.

Rawls, J. 1971. *A Theory of Justice*. Cambridge: Harvard University Press.

Rawls, J. 1993. *Political Liberalism*. New York: Columbia University Press.

Rehg, W. 1997. "Legitimacy and deliberation in epistemic conceptions of democracy: Between Habermas and Estlund." *The Modern Schoolman* 74: 355–374.

Saffon, M. P. and Urbinati, N. 2013. "Procedural Democracy, the Bulwark of Equal Liberty." *Political Theory* 41 (3): 441–481.

Schumpeter, J. A. 1950. *Capitalism, Socialism, and Democracy*. New York: Harper.

Schwartzberg, M. 2015. "Epistemic Democracy and Its Challenges." *Annual Review of Political Science* 18: 187–203.

Simmons, J. 2008. *Political Philosophy*. Oxford: Oxford University Press.

Thompson, D. 2008. "Deliberative Democratic Theory and Empirical Political Science." *Annual Review of Political Science* 11: 497–520.

Talisse, R. 2009. *Democracy and Moral Conflict*. Cambridge: Cambridge University Press.

Young, I. M. 2000. *Inclusion and Democracy*. Oxford: Oxford University Press.

Wacquant, L. 2012. "From slavery to mass incarceration: rethinking the 'race question' in the US." *The New Left Review* 13: 41–60.

Croatian Journal of Philosophy
Vol. XX, No. 60, 2020

Book Discussion

Can Statism Help?

NENAD MIŠČEVIĆ
University of Maribor, Maribor, Slovenia
Central European University, Budapest, Hungary

Can statism help with burning issues of the present time? The authors in the collection mostly answer affirmatively; in their view states can successfully deal with their cosmopolitan responsibilities. In the discussion, we question this optimistic assumption, and suggest the need for a more supra-statist, cosmopolitan arrangement for solving the issues.

Keywords: Statism, cosmopolitanism, refugee crisis, responsibility to protect.

The volume is a challenging endeavour in political philosophy. Talking about it the editors write: "By seeking to bring the state explicitly back into the cosmopolitan discourse, it helps advance enquiry into whether and how the state may be an agent of, rather than an obstacle to, cosmopolitanism" (5). Further, "(…) this book seeks to investigate the possibility that states can become bearers of cosmopolitan responsibilities while also remaining vehicles for popular self-determination within persisting, and at times counteracting, conditions of global pluralism" (5). For historical sources the editors, Garrett Wallace Brown and Samuel Jarvis, appeal to a reading of Kant

> As Kant suggests, a cosmopolitan matrix might develop from 'one powerful and enlightened nation…a republic' and that this could 'provide a focal point for federal association among other states' (Kant 1970: 104). (…) According to Kant, the motivation for joining any alliance is not determined by the 'motivations of morality', but motivated on the empirical and political realities embedded within global relations (Kant 1970: 114). (205)

The volume consists of Part I "The Responsibility to Protect as a Cosmopolitan Doctrine", Part II "Cosmopolitan Responsibility and The Legal Practice of Extraterritorial Jurisdiction", Part III "Global Issues and Responsibility Beyond the State," and the Part IV "Cosmopolitan

Republicanism." Most papers are in line with the mainstream pro-statist intention, with the exception of a few written by well-known cosmopolitans, like Luis Cabrera and David Held with which we mostly agree; we shall here focus on the mainstream papers.

Can the proposed model of independent states with some mutual interaction, but without a strong supra-state apparatus, call it "the intergovernmental model of democratic states", deal with burning crises of contemporary world, in the first line the refugee crises, and distant conflicts that have been provoking it?[1] Can its real-life incarnation protect distant people in need of protection? This is the guiding question of the present book, mostly centred upon the need to protect often distant threatened people; the first part bears the corresponding title „The Responsibility to Protect as a Cosmopolitan Doctrine," various parts and chapters reply to different sub-question. Some of these stem from global problems, like climate change (ch. 9) or health problems (ch. 10), others from specific principles governing the democratic states (the whole Part IV is dedicated to cosmopolitan republicanism). Some papers, in fact majority of them, promote a statist model, and a minority makes steps towards more cosmopolitan alternatives.

In this review, we shall stress the guiding question, and then briefly look at two additional issues, the climate change and the problem of global health. Let us then start with the paper dedicated explicitly to the crisis governing the recent political situation, Michael W. Doyle's "Global Refugee Crisis". Doyle proposes that what is required is "to reform existing global structures in order to create alternative governance pathways" (94). The pathways should reflect first 'culpability' for causing the harm, i.e. chasing away the population; he mentions examples like Daesh and Nusra. Second, it should reckon realistically with 'capability' of the candidate host countries to provide assistance, which should be, as he puts it "at least proportionate to national cost" (94).

Now, Doyle's brief proposal for how to accomplish the first task is really minimalistic. He mentions "referral" and then notes that "the UN Security Council would be justified in seizing the overseas financial assets of the Syrian state and any terrorist group with seizable assets" and using them to pay for the support of refugees on the Syrian border (92). Sounds at least minimally optimistic, but the reader learns immediately the crucial piece of information: all this would be politically very difficult "to apply in many crises" (92). Why this proposal counts as a realistic one is completely unclear to me.

Note that the wider framework of Doyle's reflection is the appeal to the "responsibility to protect" (as formulated throughout the book); the activity in question should address the causes of the refugee influx, and thus prevent future crisis of the same sort. If what Doyle is offering is all that the statist line has to offer, and we have good reasons to

[1] I am borrowing the term "the intergovernmental model of democratic states" from Michael Zürn (2018: ch. 9, passim).

assume that it is the case, it would be rational to abandon it, I would think!

Consider now the capability considerations, as presented by Doyle. He starts by reminding us of the preamble to the 1951 Refugee Convention suggesting solidarity and international cooperation. Then he reminds the reader that the application of the Convention proposal has ended with a "collapse in the EU"; he concludes that we need more modest proposals. Under the first one, countries of potential resettlement would commit to pledging the share of the identified need they will cover; under the second, "countries could make their family, labour and student visas more readily available to refugees, by giving priority to refugees and forced migrants" (94) or do something of the kind to alleviate their situation. And this very hypothetical proposal is all he offers at the side of political institution. Another way is "mobilizing the private sector" for measures like private sponsorship of refugee resettlement and the like. And this is all! No positive institutional story, and some privately organized *ad hoc* cures for the burning problems!

To conclude, let me quote Doyle's excellent diagnosis of the problem, from the beginning of the chapter.

> The current problems facing the international refugee system are deeply rooted in the dual principles of national sovereignty and universal human rights embedded in the post-Second World War global regime. The 1948 Universal Declaration of Human Rights affirms that everyone has a right to leave a country. Yet the principle of national sovereignty holds that no one has a right to enter another country without its sovereign permission. (88)

I agree about the roots of the problem, and I think it speaks against principles of natural sovereignty in favour of a more cosmopolitan arrangement. Let me stress that Doyle is one of the most competent and intelligent defenders of the *status quo* statist approach. If this is the most that can be said from the statist viewpoint, it tells a lot against the viewpoint!

The discussion of refugee crisis leads naturally to the issue of its distant causes, and prominently of violence and warfare in relatively distant regions that force people to migrate. The issue is discussed in the book relatively independently from the context of crisis; as we said it goes under the name of "The Responsibility to Protect", shortened to "R2P"; the first part of the book is specifically dedicated to it.

The first paper of the first part, by Alex J. Bellamy and Blagovesta Tacheva, entitled "R2P and the Emergence of Responsibilities Across Borders" is only partly in the spirit of the book as the whole. It does suggest, in a modest fashion that "R2P is primarily a responsibility to consider taking action to protect populations from genocide and mass atrocities—a 'responsibility to try'" (35). But when it comes to the signification of this responsibility, it turns to David Held as a guide, and Held is, of course, much less statist and more firmly cosmopolitan. Following him, our authors suggest that "although the recognition of the principle is an important and necessary first step, it should, in further

development "materialize" into a "fully-fledged cosmopolitan responsibility" (35). But for Held, such fully-fledged responsibility leads to a Global Covenant, as the title of his (2004) book suggests; and the Covenant proposes a supra-nationally controlled loose federation, very far from the statist model favored by the mainstream of the book. (We shall skip here Held's contribution to the volume, entitled "Cosmopolitanism in the Face of Gridlock in Global Governance" with which we very much agree, but which lies completely outside the mainstream project of the book).

This brings us to the last two papers of the first part, both strictly in the mainstream line. Let me start with Toni Erskine's "Coalitions of the Willing and the Shared Responsibility to Protect". As the title suggests, she proposes informal 'coalitions of the willing' as a means for discharging the obligation, and characterize them as "*ad hoc* associations". We can agree with her that states have a moral duty to establish such association. However, it is incredible that *ad hoc* voluntary associations could be the main guarantee of protection of human rights of the world population! What if we face, like we do these days, a very wide coalition of the *unwilling*, blocking informal, ad hoc attempts to the contrary? Even if we assume states to be more hospitable than they are these days, in line with authors like Kant and Seyla Benhabib, is there any force in such a system that would force, or almost force them to enter such *ad hoc* coalitions?

Things look similarly with the remaining paper in the first part, Derek Edyvane's and James Souter's "Good International Citizenship and Cosmopolitan Responsibilities to Protect".

> At its best, the ritual of parliamentary debate may also serve the function of acknowledging, containing, and communicating conflict. (…). The ritual here, of rival perspectives passionately voiced across the aisle (and on either side of it) in lengthy debate leading to a vote is of tremendous symbolic importance. But its significance is not purely symbolic. (…) the adversarial institutional framework serves both as 'an essential obstacle against the happy acceptance of the intolerable', and as the 'correct reaction' to cases of deep, and possibly ineliminable, conflict. (54)

> Specifically, we have suggested that the integrity of the good international citizen, and its commitment to cosmopolitan ideals such as R2P, is preserved by habits of reluctance and caution, its unwavering commitment to adversarial institutional frameworks, as well as by practices of reparation. So doing offers to drive a wedge, albeit a fragile one, between moral conflict and tragedy (57)

Can we really accept as a political norm a system that offers merely a "fragile wedge" between moral conflict and tragedy? In Srebrenica, a fragile wedge between the two was provided by UNPROFOR's 370 Dutchbat soldiers, and the world has learned not to trust such weak wedges. We may conclude that the upshot of the papers meant to defend the ability of a statist system to ensure rights of distant people is quite dissatisfying. However, the last part of the book offers an at-

tempt in the same direction, the difference being that it starts from a "republican" model of states offering the protection of rights and the rest of R2P repertoire.

Some papers discuss the differences between republican perspective and the alternatives (mostly classical liberal ones), stressing non-domination as the main virtue of republicanism, but without saying much about specifically cosmopolitan aspects of it. Thus, we shall skip them, and concentrate upon those that propose clear republican constraints on global actions, in particular the ones geared to responsible protection.

In her contribution entitled "The Cosmopolitan Responsibilities of Republican States" Miriam Ronzoni stresses the republican principle of non-domination, her statement most relevant for the present context, concerns republican constraints on humanitarian intervention. She notes that such interventions are heavily constrained in relation to conditions, justifying reasons and kinds of acts of war permitted. And then she adds, as the republican comment, that that "if any kind of forceful intervention entails some domination, this is always a pro-tanto evil from a republican perspective" and that "if domination can be minimized, it should" (327).

The second claim seems almost trivial: of course, if you can help the victims in a less domineering way, it would be a good thing to do. The weight of the first depends on how we read the "pro-tanto" formulation. The usual reading would take "pro-tanto" to ascribe some badness to the act in question. But take the simple analogy. Suppose John saves Jane from a serious rapist attacker by hitting him hard, and a commentator, call her Miriam, claims that the saving act was bad in a certain aspect, "pro-tanto". Commonsensically viewed, Miriam is just wrong; the act is simply not bad. A philosopher might defend Miriam by saying that "pro-tanto bad" here means just that it would have been better if there had been no need to defend the person in question, and no attack at all. True but trivial; it does not make the act of saving Jane bad in any way. The same holds for Ronzoni's view of the pro-tanto badness of a humanitarian intervention. Suppose, the attacker country A is performing genocide on members of the victim country V; the rescuer country or coalition R intervenes, and by intervening forcefully breaks the will of A leaders, thus dominating it, and saves the people of V. If Ronzoni claims that therefore the rescuer action has been bad "in some respect" she is wrong. The only badness that can be located around consist in the fact that it was bad that help was needed, like in our rapist story. But this is trivial and irrelevant for judging the act of the rescuer. (Another meaning of pro-tanto, namely *prima facie*, or defeasibly, is not relevant here: if the act is merely prima facie bad, but is in fact not bad at all, since the potential badness is defeated in the particular case, there is no disagreement about intervention. Take the example of driving through red light in order to save someone's life, as

Smiljana Gantner reminded me. Here, the possible badness is defeated by circumstances, and there is nothing bad about the action itself. Ronzoni's claim that the issues of this kind are "extremely tricky" (327) cannot hold for the reasons she is offering

An additional worry for Ronzoni comes from the following question: suppose the act that broke the will of the attackers, thus dominating them, is performed by the victims and it was dominating in minimal amount needed to save the victims. Would the act be in any respect bad? I can't see that it would.

Steven Slaughter's "Republican Citizens and Political Responsibility in a Globalizing World" discusses the ways in which political agents could "develop responsive governance beyond the state". He lists several options. "The first and most fundamental way is to pragmatically augment the virtues that inform republican citizenship" (313). His line seems to be individuals' focus. It stresses the need of presenting "cosmopolitanism as a personal virtue that informs the ways that citizens conceive of politics and direct the state" (314) reciprocal concern for all human beings as necessary to realize liberty in a highly interdependent world (see Turner 2002). Also, one should extend the range of reflective deliberation about political discourse and policies so as to encompass radical and dramatic changes like the global ecological ones.

Next, he prompts us to "rethink what contestation means within contestatory democracy to ensure that there are opportunities for deliberation between republican citizens and formal International Governmental Organizations (IGOs) set up by states. Finally, along the same lines, he notes that republican citizens should have critical respect for transnational activism and the resulting deliberations in transnational civil society. All this would help to "extend political responsibility and responsive governance beyond the state without developing a global political community or a cosmopolitan form of democracy" (316). When one looks at his 2015 book (co-authored with Daniel Bray) one wonders at how modest his new republican requirements are.

In short, the republican perspective, for all its other potential merits, seems not to offer any additional arguments for the view that a statist global arrangement can fulfil trans-national moral obligations in a stable and reliable way.

Let us now briefly move to other issues. The two that receive detailed treatment are climate change (in the chapter "Climate Change and Cosmopolitan Responsibilities" written by Helga Haflidadottir and Anthony F. Lang, Jr) and global health (addressed in Garrett Wallace Brown and Samuel Jarvis "Motivating Cosmopolitanism and the Responsibility for the Health of Others").

Haflidadottir and Lang acknowledge that things are not very rosy when it comes to the behavior of great powers in relation to climate change; they remind us that the main international document on the issue, The Paris Accord can be interpreted as a weakening of state re-

sponsibility "because it provides space for states to create their own emissions targets" (185). However, they offer an optimistic perspective on the development, claiming that "progress has happened within international legal and institutional frameworks, and that this progress is led by states" (185) A great deal of progress is due to NGOs, but some is due to activist states, and this last point is the main source of optimism. Small activist states, like Peru and Fiji lead some activism in the direction of a more just distribution of responsibilities. They conclude optimistically: "We look to progress in international law and international organizations, the realm of states, to suggest that states can and do fulfil responsibilities in regards to climate change" (199).

A more pessimistically disposed reader might wonder whether the morals of the factual story are not really opposite to what Haflidadottir and Lang propose. If the most powerful statist actors do not respect the principle of shared responsibility,[2] which is, on the contrary, promoted by non-state agencies, by supra-state organization like EU and respected by some small activist states, doesn't this point in the direction of radical non-sufficiency of the state system, and to the need for supra-statist agencies that could make the principal actors respect the principle?

Consider now the chapter by Garrett Wallace Brown and Samuel Jarvis "Motivating Cosmopolitanism and the Responsibility for the Health of Others". It comprises two parts; the first is completely general, and presents authors' contribution to the general project of the book, and the second is specifically about issues of health. The centrepiece of the first part is the idea of "transitional cosmopolitanism", characterized as a position "which sits somewhere between motivated state communal self-interest and iterative advancement towards a potential cosmopolitan condition." (214). Such cosmopolitanism would unite global cosmopolitan interests and "the self-motivated security interests of states" (214) It would rest on an "iterative foundation"; each stage S_i arrived at in a given time t_i by means M would be re-submitted to M, in order to produce the next stage S_{i+1} at the later time t_{i+1} (the shorthand is mine. Finally, "the motivation for political action promotes, in some form, a wider recognition of a common human condition that requires moral and/or political coordination and mutual responsibilities", condition "from which a form of potential cosmopolitics emerges". (214) In particular cases we need not appeal to interests of "humanity" as a whole; what is important is that the particular iterative process does enter the cosmopolitan transition.

A problem that arises at this juncture derives from the fact that the authors tell the reader nothing about the "potential cosmopolitan condition" to which the proposed stages form the transition. If the reader

[2] The authors refer to "The principle of Common but Differentiated Responsibilities and Respective Capabilities" as the valid principle that should govern the behavior of states.

consults writings by Brown (2005) and (2009) she will find out that he proposes a relatively firm kind of Kantian "world federation", a system of states with a strong cosmopolitan control: "states would have to bind themselves to additional procedures of global interdependence and to the final outcomes of a mutually agreed governance process" (2005: 518).

The second part, on global health, offers an interpretation of global health-related international documents from the perspective of transitional cosmopolitanism. Brown and Jarvis finely document the importance of the ideal of universal health coverage (UHC) contained within the Sustainable Development Goals (SDG), and adopted by World Bank in 2016. And they point out the connection to cosmopolitanism: "At least in spirit, the language of SDG 3.8 clearly captures movement towards the cosmopolitan ideals of universal care and equitable burden sharing in the distribution of health services" (218). The measures proposed, they claim, thus "meet the conditions of transitional cosmopolitanism" (218).

Let me finally mention several papers which argue that countries should use extraterritorial laws to control *the behavior of their own citizens* when abroad. For instance, Melissa Curley in her "Exporting Harmful People" notes:

> This chapter argues that Australia's use of extraterritorial law in relation to child sex tourism (CST) illustrates and supports Linklater's thesis that, for some states, globalization has concurrently led to an expanded sense of responsibility for transborder harms (Linklater 2011). The chapter demonstrates that states can incorporate cosmopolitan harm conventions into domestic legislation that serves to punish and restrict the travel of 'harmful people'—that is, those that have been convicted of sexual offences against children and are deemed likely to reoffend. (120)

I find the cosmopolitan consequences of this reasonable proposal truly minimalistic. Other papers along the same line encompass the ones by Richard Shapcott ("Cosmopolitan Extraterritoriality"), Danielle Ireland-Piper, and Andrew Linklater. We shall skip them here, with apologies.

Finally, let me say a few words about a paper that is beyond the mainstream, David Held's "Cosmopolitanism in the Face of Gridlock in Global Governance". It offers a diagnosis of the present-day problems, and a proposal of a possible pathway out of it. He starts by asserting that "global political theory (…) has reached a cosmopolitan plateau (244); cosmopolitanism is one of the main topics and standpoints in it. But we are "at a crossroads." One road leads to the rise of nationalism and authoritarianism, while another leads to a more cosmopolitan future (he compares the present situation with the one in 1930s). In his opinion, there are four reasons for this "gridlock" as he calls it: rising multipolarity, institutional inertia, harder problems, and institutional fragmentation. Each pathway can be thought of as a growing trend that embodies a specific mix of causal mechanisms. The core multilat-

eral institutions created seventy years ago, for example, the UN Security Council, have proven difficult to change (institutional inertia.) The problems on a global scale "have grown more complex, penetrating deep into domestic policies and are often extremely difficult to resolve" (250). Finally, "in many areas international institutions have proliferated with overlapping and contradictory mandates, creating a confusing fragmentation of authority" (250). I find his diagnosis very persuasive.

He proposes five possible pathways to change. First, civil society coalitions with reformist governments, second international organizations "more autonomous and adaptive", third "plurality and diversity of actors and agencies around common goals and norms" (254), fourth, possible positive consequences of "threats to major powers' core interests", and fifth, "innovative leadership as a reaction to gridlock".

To return to the mainstream papers in the book and conclude on a positive side, let me note that the papers stem from well-known authors, who really did their best to promote an up-to-date statist approach, with global ambitions. They throw interesting light on crucial problems of contemporary world, from the protection of fundamental human rights to the issues of climate change and health; it is a pity that they do not address the issues of poverty and unjust distribution in general. They suggest how far the statist international system could go in promoting cosmopolitan goals, and present an interesting challenge to the defenders of non-statist cosmopolitan alternatives geared to the realization of the same or similar goals.

References

Bray, D. and Slaughter, S. 2015. *Global Democratic Theory: A Critical Introduction*. London: Polity.

Beardsworth, R. Wallace Brown, G. and Shapcott, R. (eds.) 2019. *The State and Cosmopolitan Responsibilities*. Oxford: Oxford University Press.

Wallace Brown, G. 2005. "State Sovereignty, Federation and Kantian Cosmopolitanism." *European Journal of International Relations* 11 (4): 495–522.

Wallace Brown, G. 2009. *Grounding Cosmopolitanism: From Kant to the Idea of a Cosmopolitan Constitution*. Edinburgh: Edinburgh University Press.

Croatian Journal of Philosophy
Vol. XX, No. 60, 2020

Book Reviews

Vincent C. Müller (ed.), Philosophy and Theory of Artificial Intelligence 2017*, Springer 2018, 313 pp.*

The volume we shall review here contains over thirty articles from various authors, divided into three thematic sections: *Cognition – Reasoning – Consciousness, Computation – Intelligence – Machine Learning,* and *Ethics – Law*. In this paper we shall focus on several articles selected from the book. However, some articles will not fully feature in the review, which is due to length limitations and for which I offer my apologies to the authors.

The first section premieres with *Artificial Consciousness: From Impossibility to Multiplicity* by Chuanfei Chin, in which the author analyses various debates and viewpoints on artificial consciousness, as well as skeptical arguments such as Chalmers' Hard Problem and potential rebuttals to such arguments. The following four articles will be, with all due respect to their authors merely listed by titles: *Cognition as Embodied Morphological Computation* (Dodig-Crnkovic), *"The Action of the Brain": Machine Models and Adaptive functions in Turing and Ashby* (Greif), *An Epistemological Approach to the Symbol Grounding Problem* (Guazzini), and *An Enactive Theory of Need Satisfaction* (Human et al.).

Next article that we will delve deeper into is David Longinotti's *Agency, Qualia and Life: Connecting Mind and Body Biologically,* which presents a very interesting perspective on the Mind-Body Problem, claiming, against the Strong Artificial Intelligence thesis, that qualia are unique to biological agents and may only be produced and explained by naturally occurring neural networks. Furthermore, the article explains qualia as a form of energy, and functionally defines it as 'control signals in the regulatory processes'.

How Are Robots' Reasons for Actions Grounded? by Bryony Pierce overviews different possible kinds of grounding to conclude that a non-conscious artificial agent's actions can only ever be externally grounded in the affective responses of their users or creators, as consciousness is necessary for grounding. The author states that while a conscious robot would be capable of internal grounding, creation of such robot would not be morally permissible.

Anna Strasser's paper *Social Cognition and Artificial Agents* takes cue in increasing presence of ever-smarter technology in our daily lives and thus human society. She follows a minimal approach to socio-cognitive abilities, according to which artificial agents can be considered as having such abilities, in that they can – to a minimal extent – understand other social

agents, exchange social cues with them and show a sense of commitment. The paper certainly raises or reminds the reader of numerous questions on status of artificial agents as members of society, assuming, of course, the reader agrees with the author's conclusions.

Other articles in this first section of the book that we have not yet mentioned are *Creative AI: Music Composition Programs as an Extension of the Composer's Mind* (Moruzzi), *Artificial Brains and Hybrid Minds* (Schweizer), and *Huge, but Unnoticed, Gaps Between Current AI and Natural Intelligence* (Sloman). An article that especially deserves a separate mention is René Mogensen's *Dynamic Concept Spaces in Computational Creativity for Music*, which has, however, unfortunately proved to be too complex to separately describe in this short review.

Moving to the second section of the book, we pay some extra attention to Shlomo Danziger's *Where Intelligence Lies: Externalist and Sociolinguistic Perspectives on the Turing Test and AI.* This paper reinterprets Alan Turing's Imitation Game as a test which also includes the socio-linguistic dimension of how an agent is perceived by the society, focusing on human prejudice and our anthropocentric tendencies, which may present an obstacle in objectively evaluating artificial agents' intelligence. With this, the paper presents a rarely seen point-of-view towards human-AI relations, and, as the author concludes, teaches us 'quite a bit about human intelligence as well'.

Will Machine Learning Yield Machine Intelligence? by Carlos Zednik, published as the last article in the second section, considers how Explainable AIs might be able to solve the Black Box Problem, and what that means for the prospect of machine intelligence, concluding, after a short but concise analysis, that the answer to the question in title may well be affirmative, iff we acknowledge algorithmic similarity of AI and humans to be a sufficient criterion for machine intelligence.

Other articles in this section that merit a special mention are *Supporting Pluralism by Artificial Intelligence: Conceptualizing Epistemic Disagreements as Digital Artefacts* (Human et al.), and Yoshihiro Maruyama's paper *Quantum Pancomputationalism and Statistical Data Science: From Symbolic to Statistical AI, and to Quantum AI.*

Other articles in this section will, with all due respect to their authors, be omitted in this review to avoid the risk of ending up simply with an augmented table of contents.

The third section contains contributions on numerous ethical issues that have been subject of recent debates in AI development, such as autonomous vehicles in *Against Leben's Rawlsian Collision Algorithm for Autonomous vehicles* (Keeling), or less commonly discussed autonomous weapons in *Autonomous Weapons Systems – An Alleged Responsibility Gap* (Swoboda).

AAAI: An Argument Against Artificial Intelligence by Sander Beckers considers perils development of conscious AI presents both for humanity and such AI itself, examining the probabilities of potential conscious AIs' suffering, as well as risks a superintelligent AI could present to humans, concluding rather decisively that there should be an ethics-based ban on developing conscious artificial agents.

Another paper that caught additional attention is Abhishek Mishra's *Moral Status of Digital Agents: Acting Under Uncertainty,* which considers,

rather than the moral status of AIs as artificial agents, the moral status of agents created within simulations of such AIs. Considering grounds for moral status of such agents and the Decision Problem that arises, the author points out the numerous additional questions raised in the process of attempting to solve the Decision Problem, creating an intriguing cue for future research and discussion on the subject.

Overall, the volume offers, also within articles I have omitted in this review, for which I again extend my apologies to their authors, numerous valuable contributions to already ongoing as well as new discussions taking place on the topic of AI Theory, making it a must-read for anyone working in AI-related fields, and an intriguing mental exercise for anyone simply interested.

NIKO ŠETAR
University of Maribor, Maribor, Slovenia

Samir Okasha, Philosophy of Biology. A Very Short Introduction*, Oxford: Oxford University Press, 2019, 152 pp.*

Samir Okasha, Professor of Philosophy of Science at University of Bristol, analyses the main problems of philosophy of biology in his book *Philosophy of Biology: A Very Short Introduction* (2019), as the title suggests. Philosophy of biology is a very diverse field. Okasha presents the main problems and the current debates taking place in this discipline, highlights the key problems of philosophy of biology, and presents solutions and possible objections.

Okasha's book, despite being relatively short, addresses both topics that are central to the field, and, at the same time, other areas (e.g., molecular biology, sociobiology, etc.). At the end, the author gives excellent suggestions for further readings, arranged by chapters and topics. The book is divided in seven chapters, in which the author briefly describes the main problems and most commonly discussed topics in philosophy of biology. Some of these topics have already been mentioned in Okasha's *Philosophy of Science: A Very Short Introduction,* but in this book these topics are addressed more extensively, which means that the readers can become even more familiar with the major problems within this specific field.

Okasha draws attention to the fact that the book focuses primarily on evolutionary biology and genetics:

> The focus is mainly on evolutionary biology and genetics, as these are the areas of biology that have traditionally attracted the most philosophical interest. In recent years this situation has changed somewhat, as philosophers of biology have turned their attention to areas such as developmental biology, immunology, and microbiology. (8)

It should be emphasized, as Okasha says, that the book offers insight into philosophy of biology regardless of the readers' prior knowledge.

In the first chapter, 'Why philosophy of biology?', the author explains why philosophy of biology is important in the first place. The chapter begins with a presentation of how the study of scientific methods was transformed

into a new discipline, philosophy of science, at the beginning of the 20th century, and, as the author explains, further branched off into philosophy of biology. In doing so, the author draws on various other authors who have made important contributions to the development of the philosophy of science, such as Descartes, Leibniz, Hempel, and Kuhn.

The author then draws attention to three important factors that influenced the emergence of philosophy of biology:

> First, it become clear that traditional philosophy of science was too physics-centric—biology had been left out of the picture. Second, conceptual issues that arise within biology began to attract the interest of philosophers, leading to fruitful interdisciplinary exchanges. Third, proponents of 'naturalized' philosophy, which uses empirical science to help tackle philosophical problems, increasingly looked to biology for inspiration. (2)

At the beginning of the book, Okasha therefore commits himself to a brief historical overview of the development of philosophy of biology; this is important for understanding why biology was put in the background and why, in its initial period, most of philosophy of science was based on physics. Okasha illustrates this historical overview with examples from different fields of biology, with selected examples that support the interpretation marvellously.

In the second chapter, 'Evolution and natural selection', Okasha focuses, as the title itself tells us, on the theory of evolution and natural selection. In this chapter, he also examines the philosophical significance of this problem. Anyone dealing with the issue of natural selection and evolution cannot ignore Darwin's *On the Origin of Species*, therefore Okasha presents the essential views in Darwin's claims and briefly defines the basic features of natural selection. He includes various authors that are important for understanding Darwin's natural selection in the analysis. At the same time, he does not ignore various objections. Okasha roughly presents the difference between proximal and final questions, i.e. between how and why questions, and summarizes and describes different views on evolution.

In the third chapter, 'Function and adaptation', the author explores the concept of a biological function and gives meaningful examples that accurately outline why functionality is important in biology and why we talk about functionality at all. In this chapter, Okasha also analyses the problems of adaptation in evolutionary biology and presents similarities and differences between biological function and the function of artifacts. Okasha also analyses, as he puts it, "the most popular philosophical analysis of function-talk in biology: the aetiological theory of function. The 'aetiology' of something means its causal history" (31-32).

In the fourth chapter, 'Levels of selection', the author sets out to find answers to the problems associated with natural selection and the question of whether natural selection works on individuals, genes, and groups. Okasha presents a distinction between how natural selection can be understood: at the level of the individual or at the level of group selection. The author also deals with the analysis of altruism in connection with group selection and presents critiques of group selection. He also analyses 'kinship' selection, highlighting the differences between altruistic and selfish individuals and the associated natural selection. Okasha goes into more detail on the prob-

lem of altruism and, in connection to that, also explains Hamilton's rule, which says that altruism will develop when certain conditions (known as Hamilton's rules) are met. William D. Hamilton says

> … that altruism will evolve when a certain condition, known as Hamilton's rule, is satisfied. The rule states that rb>c, where c is the cost paid by the altruist and b is the benefit to the recipient, both measured in terms of biological fitness. The final term, r, is the 'coefficient of relationship' between altruist and recipient, which measures how closely related they are. The higher the value of r, the greater the likelihood that the recipient of the altruistic action will also possess the gene for altruism. (52)

Okasha also mentions Richard Dawkins' *The Selfish Gene*, in which Dawkins advocates a 'geno-centric' view of evolution.

In the fifth chapter, 'Species and classification', the author deals with the problem of classification in biology, which is one of the most frequently debated problems in this field. The problem itself was discussed by Okasha in his *Philosophy of Science: A Very short Introduction*, where he devoted one chapter to philosophy of biology and this topic. Nevertheless, here the author addresses the question of whether there is a right way of classifying biological species into taxon species and into higher taxa. Okasha says that "classification in science raises a deep philosophical issue" (Okasha, 2019, p. 63). Okasha presents Linnaeus' classification system, analyses taxonomy, the 'species problem' in biology, the concepts of species, the problem of species as an individual, and explains the phylogenetic system. In all of this, the focus is on authors who have had a significant impact in these areas.

In the sixth chapter, 'Genes', Okasha analyses the problem of genes and the concept of the gene. He presents a brief history of genetics and the authors that influenced this field (Mendel, Watson and Crick, etc.), and analyses the essential concepts and problems that arise in the field of genetics.

In the seventh chapter, 'Human behaviour, mind, and culture', the author deals with the connection between biology and culture in humans. He wonders if human behaviour and culture can be explained by biology. Okasha draws attention to a discussion that has been going on for some time and relates to the role of (biological) nature and environment in human development. It essentially refers to the issues surrounding the question: What makes us who we are (genes or environment; genes and environment)? Okasha gives various examples that support theses about the connection between nature and environment, for example phenylketonuria or PKU disease, "which results from a mutation that affects the ability to metabolize the amino acid phenylalanine, leading to brain damage. However, if an infant with the mutation is kept on a diet low in phenylalanine, their brain will develop normally" (105).

The author also presents the development of sociobiology, originating in the works of Edward Wilson. Okasha analyses, among other things, Wilson's examples of sociobiological explanations of social phenomena, such as incest. The author also presents critiques of the mentioned theories and at the same time explains the meaning of human culture and offers arguments for how (if at all) cultural evolution and biological (genetic) evolution are connected.

> Based on the existence of cultural differences among humans, and the fact of gene-culture dual inheritance, it is suggested that a process of cultural evolution operates alongside genetic evolution in human populations, sometimes interacting with it. (116)

The book gives a brief but very detailed introduction to the field of philosophy of biology, with the author himself emphasizing that these philosophical questions are topical, widespread, and important in biological sciences.

One can agree with Okasha's concluding note, where he states that:

> By scrutinizing the meaning of biological concepts, studying the implications of biological theories, and probing the logic of biological explanations, philosophy helps to deepen our understanding of the worldview painted by modern biology. (117)

The book *Philosophy of Biology: A Very Short Introduction* is an outstandingly interesting book and a great introduction that provides insight into essential problems touching the research area. The book is very readable and brings philosophy of biology closer to readers that do not have much experience in this field.

It is great that Okasha presented the main themes of philosophy of biology, whose issues cover several different areas and thus opens up new questions. Due to its brevity, clarity, and accuracy, the book offers the reader a solid introduction to the increasingly diverse field of philosophy of biology. The reader does not necessarily need in-depth knowledge of biology to understand it, and therefore to anyone who is interested in the topic, regardless of previous knowledge.

URŠKA MARTINC
University of Maribor, Maribor, Slovenia

Anđel Starčević, Mate Kapović, Daliborka Sarić, Jeziku je svejedno *(*Language could care less*), Zagreb: Sandorf, 2019, 376 pp.*

Jeziku je svejedno [*Language could care less*] is a book written by three young Croatian linguists from the University of Zagreb dealing with prescriptivism in general and prescriptivist practices in Croatia in particular. The topic itself is far from new. In their book *Authority in Language*, which came out in four editions (first published in 1985), James Milroy and Lesley Milroy say in the Preface that: "Essentially, *Authority in Language* explores the perennial topic of correctness in language" (xiii). And with correctness goes what is correct to use, that is prescription of correctness. They say: "If, in a particular culture at a particular time, guests at a dinner are required to wear evening dress (of a particular form) and required to use their knives and forks in a particular way, these requirements are *prescriptive*, that is, they are imposed from 'above' by 'society', not by *ad hoc* agreement amongst the guests themselves." However, they also stress that "language is a much more complex phenomenon than table manners: it is also a much more central aspect of human experience". (1)

Starčević, Kapović and Sarić's book is divided into two main parts: Part 1 "Prescriptivism and the ideology of the standard language" is of much more interest to general linguists, psycholinguists, cognitive science prac-

titioners and especially political philosophers since it is more theoretically oriented and puts stress on basic findings on communication, language variability, standard/nonstandard varieties of language and focuses on the important questions of language policy and language planning. Part 2 is a scathing critique with many vivid and damning examples of Croatian prescriptivists' catastrophic practices, and in many cases even more ludicrous usage advice they give to ordinary speakers. For the audience of this journal I shall concentrate more on the first part of the book and leave the second to the delightful inspection of Slavic linguists.

How do the authors characterize prescriptivism? It is a conservative language ideology and they draw a strong parallel with any other (political) conservative ideology: there is an insistence of the *status quo,* so called stability (language as it is). There is resistance to change. Language change is seen as deterioration or, in Jean Aitchison's words, language decay. There should be order, the standard language should obey strict language norms, that is the rules of the standard language (which the authors insist on calling the standard dialect), tradition, obedience of authority (which leads to a proliferation of language manuals, usage guides, dictionaries, the glorification of the speech of educated people, "good" writers, etc.). This includes the promotion of national unity (under the banner of the standard/national dialect), resistance to the so-called anarchy of spoken language and also to the dialect/language contact. They, I think rightly, believe that such characterizations are easily recognized as features of conservative ideology in general (134).

It is believed (but not uniformly) that prescriptivism is not part of linguistics and that prescriptivists are not linguists since they propagate attitudes that are not part of scientific linguistics. Linguists describe, they do not prescribe. Prescriptivism has been labeled 'amateurish linguistics'. It is interesting, and partly amusing, how Steven Pinker refers to prescriptivists in his most recent usage guide, (aimed at improving the style of good writers), *The Sense of Style: The Thinking Person's Guide to Writing in the 21st Century*. He wonders: "Who are these writers? You might think I'm referring to Twittering teenagers or Facebooking freshmen. But the writers I have in mind are the purists—also known as sticklers, pedants, peevers, snobs, snoots, nitpickers, traditionalists, language police, usage nannies, grammar Nazis, and the Gotcha! Gang… The idea that there are exactly two approaches to usage—all the traditional rules must be followed, or else anything goes—is the sticklers' founding myth" (323–324).

People attach great value to many different phenomena and thus also to language in particular. As the Milroys stress: "Ordinary people (i.e. nonlinguists), however, have been accustomed from time immemorial to make value judgments about language" (1991: 10). In a most recent book on prescriptivism by Ingrid Tieken-Boon van Ostade, *Usage Guides and Usage Problems in British and American English* (Routledge 2020), many pages are devoted to the explanation of these bottom-up prescriptive efforts of ordinary language users from all social backgrounds, which are referred to (by Morana Lukač) as 'grassroots prescriptivists'. Bottom-up or grassroots prescriptive efforts are understood as those initiated by lay members of the general public. The book under review is not about them but about top-

down prescriptivism, which is mostly carried out by so-called language authorities. The book is actually a defense of ordinary people's inquiries about language from, in Pinker's words, "pedants" or "language police".

As other (mostly Anglo-American) linguists have done, the authors of this book show where prescriptivists or purists have gone wrong. Unacceptable are subjective proclamations about right and wrong in language, the talk about better or worse language forms. Purists have a strong inclination to select one, and only one, from a set of other, equivalent forms and usages and recommend that one as the 'correct' form. But, as it is well known, no language or dialect can be shown to be better or worse than another *on linguistic grounds alone*. The authors especially stress that language is rarely simple, binary, black and white, without exceptions, formally consistent and symmetrical. Language as a social fact is similar to its speakers – complex, multicolored, multifarious, full of unusual forms, meanings, expressions and usages. They firmly assert: "Language is simply like that and it could not be different" (75).

The authors clearly say and argue that the conservative view of language as advocated by purists/prescriptivists goes hand in hand with conservative right-wing politics. It has been noticed and stressed that although discrimination on the grounds of race, religion, gender and social class is not publicly acceptable any more, it appears that discrimination on linguistic grounds *is* publicly acceptable (Milroy and Milroy 1999: 2). Thus the duty of linguists is to react critically if they notice that certain individuals in positions of power and supposedly in the name of linguistics manipulate and purposefully misguide or wrongly inform the ordinary speakers about linguistic issues and thus instigate collective language insecurity and create and perpetuate social inequality (68). The authors stress that language policy, and with it language planning, is primarily politics so it is unacceptable to present language planning as an objective, scientific and neutral activity that creates a more perfect, more precise and more economical language variety, i. e. the standard variety. What is in fact done is the ideological act of persecution of other language forms and varieties. Furthermore, there is no language without variant forms of some words, no language with one meaning per word and no language without mixing of codes and registers. Why should the standard language be as remote as possible from actual usage? Why is the standard better when it has less to do with ordinary, i.e. real language use? (188). What is the use of the artificiality of the standard? The authors think that this is no service to speakers but it serves very well the purpose of authors of usage guides and their numerous editions, which brings them a substantial profit.

What is the result or consequence of an uncritical acceptance of usage guides? First, they create language insecurity so that ordinary speakers feel that they do not know their own language. In a vivid metaphor by the journalist Jurica Pavičić "speaking and writing one's mother tongue becomes like walking on a minefield in which any wrong move/step can the fatal" (*Jutarnji list*, a Croatian daily newspaper). The second consequence is the manipulation of the average speaker. In Einar Haugen's terminology, it creates language schizoglossia, language insecurity where the speaker is ashamed of using his/her own language and where s/he develops a fear of

his/her native language. A particular political reference is to the 1990s in Croatia, when people, in order not to be accused of writing in Serbian, used newly and artificially created Croatian words, many of which fortunately did not survive. Thirdly, prescriptivists spread out the idea that language is attacked by foreign (today mostly English) words, more particularly Serbianisms. Fourthly, a widespread belief that language is decaying and not simply changing. This is, of course, the idea held by most ordinary speakers, who think that language in general was better in earlier generations. This prevalent, yet inaccurate, belief of common speakers that language was once at its golden peak is further used and manipulated by prescriptivists to propagate the view that, for example, foreign words 'weaken the resistance /hardiness of national being' and that they 'cloud the real meanings'. The authors rightly conclude that "this is an illusion but a persistent implicit or explicit prescriptivist mantra which induces in speakers an unfounded feeling of fear about a pending language chaos and inability to express thoughts" (82).

The authors have been accused that in their book they advocate communicative chaos and that they are against the standard language. Communicative chaos is a non-existing phenomenon and the authors have stated many times (especially in their interviews and presentations of the book around Croatia) that they are not against standard (which would be absurd to start with) but that they are against rigid and crude prescriptivism. In their own words, talking about the aim of the book they say: "In this book we deal not only with the deconstruction of the mystificatory nature of prescriptivism but we are also concerned with exaggerations in the normative practice/activity, primarily in different kinds of usage guide books and also in the choice of elements/forms that prescriptivists want to change, mould or throw out/dispose of. Moreover, we are concerned with their non-scholarly argumentation that goes with such advice, based on a non-scholarly interpretation of linguistic phenomena and thus inconsistent/incoherent implementation of language 'corrections'" (127).

The book is written in a clear argumentative style with a lot of vivid comparisons of which I will mention only a few. ….*While the astrologer does not claim to be an astronomer, the prescriptivist pretends to be a linguist (43). Should the standard language be like a Greek vase in the museum, the artifact that should be admired, around which one should walk gingerly as if on eggshells and not to be used on any occasion since it might get damaged? (189).* In their interviews they say: Prescriptivists sell *poisonous fog*, the myth about language chaos serves to *frighten little children*. To give prescriptivism a scientific status would be *as if the physicist advocated bilocation* or if a medical doctor said that *in hospitals priests are more relevant than doctors*.

What they advocate for language is directly correlated to their political beliefs. We are openly political, they say, and their criticism of prescriptivism as scientifically unfounded goes hand in hand with their justified belief that prescriptivism is politically backward and that produces damaging social consequences. Prescriptivism goes with conservative language ideology and as we have conservativism (right-wing) in politics, we see all of that reflected in prescriptivists' attitudes to language. Just as conservativism is tied to nationalism so is prescriptivism.

In conclusion, this book is worth wider public notice of linguists, cognitive scientists and in particular political philosophers interested in language. Here are three reasons for it, the most important mentioned last: 1. It is the first thorough criticism of prescriptivist practice in Croatia, 2. It is a valuable addition to numerous books on language prescriptivism in general and 3. Its main stress is on the political underpinnings and moreover and more importantly, political repercussions for wider society and its speakers that prescriptivists bring about by their persistent and unnecessary advocacy for language corrections.

Linguistic activism or critical linguistics is the activity of linguists who fight against using language as an instrument for social discrimination and manipulation. This book is a first-rate act of activism. Their fight against language discrimination, against a rigid standard language and for its spontaneous change and for a legitimate recognition of various dialects goes hand in hand with their belief that nobody should be discriminated socially, e.g. because they are poor or because they are female or homosexual. Jurica Pavičić, the above-mentioned writer and journalist, has said that this book, apart from being a highly relevant scientific book, is *politically perhaps the most important book* that appeared in Croatia in 2019. Behind the authors' liberal views on language policy and language planning lies hopefully a more liberal Croatia.

DUNJA JUTRONIĆ
University of Split, Split, Croatia

Larry Krasnoff, Nuria Sánchez Madrid, Paula Satne (eds.), *Kant's Doctrine of Right in the Twenty-first Century,* Cardiff: University of Wales Press, 2018, 304 pp.

Kant's legal and political philosophy has often been viewed as his least important contribution, branded by the likes of Hannah Arendt as a work of his autumn years, and vastly overshadowed by his contributions to ethics, epistemology, and even aesthetics. The editors of this volume have sought to remedy that by showing that the renewal of interest in that part of Kant's work, visible in the last two decades, is in fact justified by the potential inherent in it. This collection of articles, addressing the topics of Kant's legal and political *magnum opus*, his *Doctrine of Right* (published in 1797 as the first part of *The Metaphysics of Morals*), has thus two main, interwoven ambitions. The first is to show that his legal and political philosophy comprises a domain separate from his ethics, interesting and fruitful in its own right. The second is to suggest some of the ways in which that autonomous domain "might have real application[s] to the contemporary world" (*Introduction*, 1)

The twelve articles in the volume are obviously not intended to provide a comprehensive guide to the whole *Doctrine of Right*, but they manage to cover its main topics taken as potential contributions to contemporary debates. These include the nature and the role of the social contract, the justifiability and content of human rights, the purpose of the state (especially in the context of debates about its role in the welfare of its citizens), the limits of political authority and obligation, international relations and, finally, in-

terpersonal relations viewed with the benefit of distinction between ethical and juridical domains. All of the important topics are addressed, although some are approached from unusual perspectives, which is definitely a conceptual plus, because it adds to the argument of the fruitfulness of applying Kant's insights to contemporary debates. Opposing views are also included, which Kant himself, as a staunch supporter of public intellectual debates, would surely have welcomed.

The opening chapter of the book is fittingly selected for its attempt to present Right as a domain independent of ethics, establishing it as a standalone part of Kant's moral philosophy. Concerned only with duties for which external lawgiving is possible, whose enforcement can be secured by lawful coercion, and thus only with external freedom, Right is clearly distinguished from ethical considerations, while at the same time included under the obligations dictated by practical reason. Starting with that demarcation, Macarena Marey then continues by establishing the role of the social contract in Kant's philosophy. It is used as a normative idea, intended to provide the justification for the state's monopoly on the legitimate use of coercion, at the same time establishing the state as on obligatory end in itself, necessary for achieving external freedom and equality. It is not used to legitimize ethical norms, nor is it used in purely pragmatical reasoning, aimed at ends such as the maximization of happiness. Marey thus suggests that Kant's purely juridical standpoint is a welcome addition to contemporary debates concerning the nature and the role of the social contract, which have been revitalized ever since Rawls renewed the interest in the subject.

Addressing the same topic, Alice Pinheiro Walla aims to show that the social contract is not seen by Kant exclusively as a normative standard, nor as a useful heuristic principle. Instead, she claims that the united general will, seen as an entity established by the actual consent of all human beings, is a real goal envisioned by Kant, necessitated by the need to establish legitimate private property claims. Without it, the use of external objects by finite rational beings can never be fully compatible with the external freedom of all, thus making all property rights prior to establishing a global (republican) political community merely provisional.

The interconnected topics of human rights and the purpose of the state take up a third of the book. In an attempt to show the way in which Kant's insights might bring something new to the table, Eric Boot starts with the observation that in current liberal theories of human rights there is a tendency to treat rights as a fundamental moral category, then used to develop the corresponding duties supposedly based on them. Kant's approach turns things the other way around. Duty is the fundamental moral category, and rights are established only on the basis of those duties which Kant calls perfect duties. Their "perfection" stems from the fact that they prescribe with precision who owes what to whom, thus making the enforcement of them possible, in principle, by lawful coercion, which makes them juridical duties (the only exceptions being the duties not to lie, not to commit suicide and the duty of respect, which are perfect but concern internal motives, not external actions, thus making them the only perfect ethical duties). The upshot of Boot's observation could be the potential to stop the hyper-inflation (and the corresponding devaluation) of human rights. The downside, at least for

some, is the fact that Kant sees the duty of beneficence as an imperfect duty, its imperfection stemming form its latitude, which does not prescribe with precision who owes what to whom, implying that there can be no such thing as a right to, for instance, a decent standard of living.

Picking up where Boot left off, Masataka Oki tackles the question of the purpose and the role of the state, pointing to the obvious fact that Kant does not see the state as an entity tasked with promoting the material well-being of its citizens. This point connects with Boot's claim of the impossibility of establishing the normative foundations of the modern welfare state based on Kant's legal and political philosophy. However, Oki claims that the functioning of the Kantian state is internally connected with human happiness, but in a different way. Man, being the sensual, finite but also rational being that he is, requires freedom for his happiness. And human freedom, Kant claims, is only possible within the confines of a civil condition, in which the external freedom of all is made possible by universal laws, made by all (as lawgivers), and applied equally to all (as citizens).

Driving the point further, Nuria Sánchez Madrid acknowledges Kant's claim that the state can and should tax the wealthy citizens to support those who are destitute, but points out that claims like that in no way support the reading of Kant as an advocate of the welfare state. Proceeding not from a supposed right to a decent standard of living, but from the state's duty to preserve the civil condition (which might be endangered by internal strife, external aggression prompted by the state's weakness, or simply by losing parts that make up the whole), she claims that these measures simply don't equate to the current concept of poverty removal as a duty on national as well as global levels. Limited in their scope to the preservation of life, in Kant's vision they do not include the removal of social inequality, which is seen as compatible with republican citizenship and equality before the law.

A dissenting voice on the topic can be heard from Larry Krasnoff, whose thesis is that the *Doctrine of Right* should be read as a rejection of the familiar difference between classical and welfare liberalism altogether. The supposed difference rests on the premise that law and freedom can be separated. The so called classical liberals maintain that freedom predates law, which can then be seen as a potential threat to freedom, and that the state should be concerned only with the legal equality of its citizens, rejecting modern welfare programs as attacks on individual liberty. The so called welfare liberals maintain that law in a way creates freedom, making it conditional, and enabling the state to pursue welfare programs after freedom has been secured; those programs are then legitimized by considerations other than political liberalism (usually by a version of utilitarianism). In the *Doctrine of Right*, on the other hand, Kant claims that (external) freedom and law are essentially connected. Freedom cannot be seen as predating positive law, nor can positive law be normatively conceptualized without the concept of mutual external freedom (thus it cannot be said that we start with positive law, and only then come to understand what freedom is). From there, Krasnoff builds up an argument that modern welfare programs can be legitimate from a liberal perspective, but not as contributions to some further goal beyond freedom (such as material well-being), but as freedom-enabling devices, making individuals independent of the choice of others.

A pair of articles in the mid-section of the book tackles Kant's infamous claim that resistance to political authority can never be justified, a claim based on the very logic of sovereignty, which precludes the possibility of legitimate public action bypassing the sovereign (and made worse by the usual interpretation that Kant considers all existing regimes as legitimate, no matter how far they stray from his proclaimed republican ideal). Wendy Brockie delivers an overview-style text, pointing to the key aspects of the topic (whether all regimes can be considered legitimate, passive vs. active resistance, the role of free speech in peaceful reform measures envisioned by Kant). Also pointing to some contemporary contributors to the debate, she concludes by observing that there seems to be no ground in the *Doctrine of Right* that can legitimize resistance to an abusive regime. Alyssa R. Bernstein offers an alternative take on the subject, pointing that there is room in Kant's political philosophy for the claim that not every thug wielding organized power must *eo ipso* be considered a sovereign, commanding respect in a civil condition. Contributing original content to the debate, she proceeds to describe a scenario in which there is even a Kantian basis for acts of civil disobedience in a legitimate civil condition, provided some very specific circumstances are in place.

On the other end of the popularity scale, Kant's most beloved politically-related claim, and arguably his most influential contribution to political philosophy to date, is his insistence that perpetual peace and global political community are to be considered the highest political good, and thus pursued as the obligatory final end of politics. The two articles covering the subject here aim to avoid the head-on tackling of the obvious points of discussion, which have been extensively covered in the existing literature (a global state vs. a global confederacy; the question of legitimate means of establishing a global community). Approaching the subject of legitimate means from a different angle, Milla Emilia Vaha points to an implication of Kant's theory that modern liberal authors, inclined to draw inspiration from Kant, might not welcome. Observing their tendency to ascribe the full extent of rights in the international domain to liberal-democratic states only, she claims that Kant's theory does not support such a view. Stemming from the moral personhood of the state, equal rights and duties belong to all states alike, even non-liberal ones (or, in Kant's parlance, despotic ones, which nevertheless are full-blown states). A consequence of this approach is that liberal states cannot have a right to meddle in the internal affairs of non-liberal states (although the question of whether all coercion-enforcing regimes qualify as states remains open).

Sorin Baiasu makes his contribution to the topic even more original, approaching the question of the means from an epistemological perspective. Claiming that there are important differences between Kant's highest ethical good and his highest political good (again delineating juridical and ethical domains of Kant's moral philosophy), Baiasu sets out to differentiate their respective guarantees. Opting for the interpretation in which perpetual peace is seen as secured by the outward workings of nature (irrespective of the internal human motives), Baisu points that its guarantee is then the object of a *doctrinal* belief, centered on the teleological picture of nature, as opposed to a different epistemic category, a *moral* belief in the postulates of

pure practical reason (god, freedom, immortality), which is needed for the possibility of attaining the highest ethical good.

Closing the volume are two articles that approach the topic of interpersonal relations, suggesting some of the ways in which their ethical and juridical aspects should be delineated. Paula Satne tackles the question of punishment and forgiveness in Kant, elegantly bypassing Kant's second most notorious legal claim, his endorsement of capital punishment. Instead, she focuses on the category of punishment in general, describing that an individual, in principle, cannot administer punishment for the wrongs done to him. Transgressions of positive law can only be punished by the state (because unilateral use of force can never be in accordance with right), while transgression of moral law can only be punished by god (the one who hands out happiness in proportion with one's worthiness, which requires being able to see a person's internal motives). On the other hand, forgiveness is a strictly ethical concept, playing an important role in a person's moral development, but having no place in the juridical domain. Forgiving a transgression of positive law would constitute a breach of the universal principle of right, by putting one person's external freedom above another's, even if the victims are willing to forgive the perpetrator who shows true remorse.

Turning from vengeance to passion and lust, Jordan Pascoe tries to evaluate the potential contribution of Kant's claims about marriage to contemporary debates about the purpose and accessibility of marriage in general. Originally conceived by Kant as a juridical solution to an ethical problem (how to enable sexual relations without at the same time debasing persons by treating them as mere instruments of pleasure), marriage becomes a part of private right. It creates a special juridical domain, in which married persons share common ends and purposes, thus precluding them from treating each other as mere means for the gain of personal ends. At the same time, the existence of their common ends and purposes is acknowledged by the state, which is thus unable to distinguish between the partners, in turn giving them both equal standing before the law. Kant thus treats marriage in the same manner that he treats private property and contracts - he sees them as legal means that enable finite rational beings to satisfy their natural needs in a way consistent with the dictates of pure practical reason. Pascoe claims that such a view on marriage has some potential to contribute to the contemporary marriage equality movement, but is not useful in more radical attempts to transform the very concept of marriage in the name of promoting social, economic and gender equality.

In summary, the volume does achieve the goal of offering a peek into "what a political position grounded in the *Doctrine of Right* would look like in twenty-first-century terms" (2). The twelve articles, obviously, do not offer a comprehensive analysis of the whole *Doctrine of Right*, but nor were they supposed to. They do give the reader a taste of Kant's main legal and political preoccupations, and suggest some of the ways in which his arguments could be brought to bear on contemporary issues. Considering the nature of the source material, a twelve-article volume could hardly be tasked with more than that.

LOVRO GRGIĆ
University of Rijeka, Rijeka, Croatia

Béatrice Longuenesse, *I, Me, Mine: Back to Kant and Back Again*, Oxford: Oxford University Press, 2017, xix + 257 pp.

In her rigorous yet pellucid reading of Kant, Béatrice Longuenesse draws her attention towards the conceptual role of the first-person 'I' in 'I think'. The concept and word 'I' is the bedrock from which all sensations, feelings, and emotions become propositionally self-ascribed. Pace Kant, Longuenesse makes the stipulation that there is a kind of consciousness propping up our ability to make first-person judgments that is *more fundamental* than the consciousness of our body. This means that the ability to recognize and reference oneself as the 'I' engaged in cognizing is more conceptually fundamental than being an embodied, spatiotemporal entity. *I, Me, Mine* offers an entirely novel reconstruction of Kant's apperceptive 'I think' before naturalizing the 'I' in Kant via Freud's developmental concept of the 'ego' (*das Ich*), which, vis-à-vis metapsychology, engages with the structural organization of mental processes qua emotion. Longuenesse ultimately draws out how Freud's 'Ich' and its developmental account bears a striking similarity to the unity of representational contents that determines Kant's 'I'.

In reading Kant, Longuenesse's metaphysical stipulation is that consciousness in the rational unity of our thinking is *more fundamental* than consciousness of our proprioceptive body, for being attentive to the rational unity of content(s) in one's thinking is what *makes it possible* to assess the standpoints from which we initially formulate, and then arrive at, shared universal conclusions. The quilting point of Longuenesse's project is as follows: the availability of the concept 'I', as the concept referring, in any instance of its use, to the entity of which the predicate in the proposition currently thought, 'I am F,' is true—that is, the availability of the concept 'I' presupposes the capacity to think writ large (83). It follows that thinking means unifying and articulating the contents of mental states into concepts, propositions, and inferential patterns. For Kant, there is a fundamental difference between the self-consciousness proper to the thinking subject in the course of their thinking and the consciousness of themselves as an object in the world. This difference is made even more distinct if, in the latter instance, we mean one's consciousness of themselves as an embodied entity—the 'I' as a physical object. It follows that what Kant means by *consciousness of oneself as a thinking subject* is not and cannot be reduced to consciousness of oneself as a physical entity, as the philosopher Quassim Cassam claims, despite it *is* intimately connected with the consciousness one has of one's own body (viz. proprioception). Just as Cassam has argued that awareness of oneself as a physical object is necessary to ground self-consciousness, Gareth Evans has proposed that the body is necessary to furnish any referential use of 'I,' including the self-ascription of mental states. Longuenesse's intervention is, thus, to stake fertile grounds and oppose such readings, illuminating a more fundamental conceptual role that the first-person 'I' expresses.

Consequently, Longuenesse begins the book by opposing this recent position amongst Kantian interlocutors, a tradition that attempts to map Wittgenstein's use of 'I' as subject onto Kant's consciousness of oneself as

subject. Longuenesse takes issue with those who emphasize the body as grounding any self-referential use of 'I', where our having available a concept of ourselves as a physical thing located in space is necessarily involved even in the self-ascription of beliefs and experiences—namely, in the self-ascription of mental predicates. Longuenesse thinks that this view, where embodiment is the bedrock for any use of 'I', meets the proper epistemic demands for the *usage* of 'I' but does not capture the semantics of 'I'. Longuenesse makes the case that Kant's "representation with consciousness," which posits the mental ascription of identities and differences, is a kind of higher-order consciousness that depends on phenomenal consciousness—the qualitative 'what it's like' to be the thinker of 'I' (i.e., 'I' as subject). According to Longuenesse, this is already in Kant's Transcendental Unity of Apperception (TUA), which is governed by the rules of *imagination* as well as logical rules of judgment and inference according to which the contents of one's representational states are reliably related to independently existing objects. Whether 'I' is a use of 'I' as subject or 'I' as object, Kant's TUA enumerates a common ground: "the unity of self-consciousness … makes possible both our synthesizing representations into conceptualizable wholes … and our ascribing thoughts to ourselves in the proposition 'I think'" (31). Thus, *all* uses of 'I,' even uses of 'I' as object, depend at least in part on the kind of information on which the uses of 'I' as subject depend vis-à-vis the *activity of thinking*—what is available to subjects via the combination of representations, which transpires as the "what-it's-like-for-the-subject-of-thinking", or the "mode of presentation" of the "'I' as subject" (31). One's qualitative awareness of thinking is *immune to error through misidentification* as the fundamental reference rule (FRR) for 'I' steps in: "'I' is a word or concept that refers, in any instance of its use, to the author of the thought or the speaker of the sentence in which 'I' is being used" (23). Longuenesse's intervention stresses that the concept 'I' presupposes the exercise of the capacity for unifying and conceptualizing mental contents: the very exercise of this capacity is conceptually expressed in the proposition 'I think'. All other judgments are subsidiary.

Satisfying the FRR for 'I' calls for nothing more than being the thinker of the thought and speaker of the sentence in which 'I' is used. Having available the fact that one is, in any given instance of one's use of 'I', the entity satisfying its reference rule requires the *awareness of one's being engaged in thinking*. Yet, in many uses of 'I', the predicate that is self-ascribed is a predicate referring to some *bodily property*. Reviewing case studies of deafferented patients and referencing Oliver Sacks' research, Longuenesse turns to empirical examples to demonstrate that proprioceptive consciousness of one's own body follows from a more fundamental consciousness of oneself *as a self*, an entity that counts as the referent of 'I' whenever 'I' is used by that entity (33–34). Longuenesse argues against the view that it is a *necessary condition* on the very possibility of a referential role for 'I' that its referent be an embodied, spatiotemporal entity, and that all 'I'-users be aware of themselves as such an entity. It follows that consciousness of oneself as thinking is, as a matter of empirical fact rather than as a matter of *a priori* argument, intimately connected to awareness of one's own body.

Longuenesse then turns to Kant's view of self-consciousness by focusing on Kant's Transcendental Deduction of the Categories and Paralogisms of

Pure Reason. Longuenesse first deracinates the 'I think' from the TUA by analyzing Kant's three syntheses: the presentation of sensory information, (its) reproduction in imagination, and apperception:

> For 'recognition under a concept' to occur, all past representations, their regularly occurring patterns, the reproductive associations those patterns have elicited, all those features must remain available for recognition. In other words, they must remain available for use in one and the same activity of apprehending, reproducing (according to associative rules), and recognizing. This is how concepts, and thus representations of objects as falling under those concepts, are eventually acquired. (79)

As the precondition for objective representation of the world, 'I think' indexes the unity of mental activity that conditions all particular instances of 'I think P'. Parsing the Transcendental Deduction, Longuenesse shows how Kant develops an analysis of the type of self-consciousness grounding the proposition 'I think' and the role of 'I' in that proposition. Countering Descartes' cogito argument, Kant's Paralogism of Substantiality criticizes the fallacious inference by which rationalist metaphysicians support their claim that the referent of 'I' in 'I think' is a soul, a thinking substance. In the Paralogism of Substantiality, Kant criticizes the inference by which rationalist metaphysicians support their claim that the soul is also a simple substance, distinct from the body. Developing these paralogisms, Longuenesse underscores that the concept 'I' is but a logically singular term, expressing the thinker's consciousness of the logical connectedness of their thoughts.

In the Paralogism of Personality, Kant criticizes rationalists for, again, engaging in a fallacy of equivocation. For rationalist metaphysics, as thinking beings we are immediately aware of our own 'numerical identity' at different times and, as such, we are persons. Kant argues that we are aware of our own 'numerical identity' at different times not in virtue of mere thinking (thinking 'I think'), but insofar as we, as thinking beings, are capable of consciousness of our continued existence as spatiotemporal, empirically given, embodied entities. Accordingly, Longuenesse develops a positive notion of persons as embodied entities endowed with unity of apperception. This is not the conclusion in Kant but one Longuenesse teases out. Longuenesse denies that the consciousness of the 'numerical identity' of any entity, including oneself, is possible for us other than by relying on criteria for identifying and re-identifying that entity in space and time.

The negative result from Kant's Paralogisms is that even though, in thinking, we develop an implicit or explicit conception of ourselves as the agent of our thoughts—indivisibly present in all instances of our thinking, numerically identical in different times and distinct from our bodies—that conception has neither *a priori* metaphysical support nor empirical support. We cannot derive any objectively justified belief in our persisting existence from the *first-person* consciousness of ourselves in thinking. Hence the positive thesis: the only way we are objectively justified in believing ourselves to be entities that persist through time, and the only way we are able to track our own existence through time, is by adopting a (supplementary) *third-person standpoint* on our own existence as the existence of an embodied entity.

Longuenesse then scrutinizes Kant's Third Antinomy and the concept of "person" as an "empirically accessible" entity that is not "necessary and

sufficient for practical use" (152). As one cannot posit an "uncaused cause" *sui generis*, the *ratio cognoscendi* of metaphysical freedom must include not only a psychological notion but also a moral notion. That is, a person is a conscious being that has a *rational will*, a faculty of desire determined under moral law—this is linked to the "moral 'I ought to'"(173).

Longuenesse argues that Kant's view of the structure of our mental life, grounding the use of 'I' in 'I think' and in 'I ought to', finds a descendant in Freuds 'ego' and 'super-ego.' Notably, Freud underscores that Kant's Categorical Imperative is the "'direct heir of the Oedipus complex'" (220). Unlike the unity of apperception expressed in 'I think', indexed to a particular body and extending to self-consciousness in instrumental and prudential reasoning, the use of 'I' in the moral command depends on consciousness of oneself as the subject of an activity of reasoning that determines the maxims of one's actions under the Categorical Imperative's unconditional demand (217). Freud's account of the structure of mental life—'ego'—provides a developmental story for the unity of apperception grounding the use of 'I' in 'I think'; similarly, Freud's account of the unconscious component—'super-ego'—and its compulsive power provides a developmental story for the conflicted structure of mental life that grounds Kant's use of 'I' in the moral 'I ought to'. The unconditional character of morality is, for Kant, originally grounded in pure reason affecting the faculty of desire; for Freud, it is originally grounded in the raw emotion that binds us to the (authority-) figure "from which we have learned the rules of our socialization" (221).

Longuenesse's project defends the intimate connection, and distinction, between *consciousness of oneself in thinking* and *consciousness of one's own embodied existence* without appealing to the noumenal realm. Specifically, Freud's metapsychological analysis of mind naturalizes Kant's analysis of 'I' in its theoretical and practical uses: 'I think' and 'I (morally) ought to', respectively. Longuenesse thereby outlines a kind of self-consciousness that, while intimately connected to consciousness of one's own body, is nevertheless distinct from it and is, moreover, *the condition for any use of* 'I'. Bridging Kant and Freud while unspooling an intervention countering those readers of Kant who have recently set embodiment as the fundamental aperture for cognition, Longuenesse's comprehensive project sets her beside those philosophical giants like P.F. Strawson and Wilfrid Sellars who have prodded forth novel modes of naturalized Kantianism.

EKIN ERKAN
CUNY Graduate Center, New York, USA