

CROATIAN JOURNAL OF PHILOSOPHY

Articles

Lexical Flexibility, Natural Language, and Ontology
CHRISTOPHER A. VOGEL

Socio-historical Causal Descriptivism.
A Hybrid and Alternative Theory of Names
CHEN BO

Becker, Ramsey, and Hi-world Semantics.
Toward a Unified Account of Conditionals
CHENG-CHIH TSAI

Plural Voting and J. S. Mill's Account of Democratic Legitimacy
IVAN CEROVAC

Is The Standard Definition of Knowledge Incomplete?
ANGUEL S. STEFANOV

Book Discussion

Consumer Complicity and Labor Exploitation
GILLIAN BROCK

Croatian Journal of Philosophy

1333-1108 (Print)

1847-6139 (Online)

Editor:

Nenad Miščević (University of Maribor)

Advisory Editor:

Dunja Jutronić (University of Maribor)

Managing Editor:

Tvrtko Jolić (Institute of Philosophy, Zagreb)

Editorial board:

Stipe Kutleša (Institute of Philosophy, Zagreb),

Davor Pećnjak (Institute of Philosophy, Zagreb)

Joško Žanić (University of Zadar)

Advisory Board:

Elvio Baccharini (Rijeka), Carla Bagnoli (Milwaukee),

Boran Berčić (Rijeka), István M. Bodnár (Budapest),

Vanda Božičević (New York), Sergio Cremaschi (Vercelli),

Michael Devitt (New York), Peter Gärdenfors (Lund),

János Kis (Budapest), Friderik Klampfer (Maribor),

Željko Loparić (Sao Paolo), Miomir Matulović (Rijeka),

Snježana Prijić-Samaržija (Rijeka), Igor Primorac (Mel-

bourne), Howard Robinson (Budapest), Nenad Smokrović

(Rijeka), Danilo Šuster (Maribor)

Co-published by

“Kruzak d.o.o.”

Naserov trg 6, 10020 Zagreb, Croatia

fax: + 385 1 65 90 416, e-mail: kruzak@kruzak.hr

www.kruzak.hr

and

Institute of Philosophy

Ulica grada Vukovara 54/IV, 10000 Zagreb, Croatia

fax: + 385 1 61 50 338, e-mail: filozof@ifzg.hr

www.ifzg.hr

Available online at <http://www.ceeol.com> and www.pdcnet.org

CROATIAN
JOURNAL
OF PHILOSOPHY

Vol. XVI · No. 46 · 2016

Articles

- Lexical Flexibility,
Natural Language, and Ontology
CHRISTOPHER A. VOGEL 1
- Socio-historical Causal Descriptivism.
A Hybrid and Alternative Theory of Names
CHEN BO 45
- Becker, Ramsey, and Hi-world Semantics.
Toward a Unified Account of Conditionals
CHENG-CHIH TSAI 69
- Plural Voting and J. S. Mill's Account
of Democratic Legitimacy
IVAN CEROVAC 91
- Is The Standard Definition of Knowledge Incomplete?
ANGUEL S. STEFANOV 107

Book Discussion

- Consumer Complicity and Labor Exploitation
GILLIAN BROCK 113

Lexical Flexibility, Natural Language, and Ontology

CHRISTOPHER A. VOGEL*

Department of Philosophy, University of Maryland, USA

The Realist that investigates questions of ontology by appeal to the quantificational structure of language assumes that the semantics for the privileged language of ontology is externalist. I argue that such a language cannot be (some variant of) a natural language, as some Realists propose. The flexibility exhibited by natural language expressions noted by Chomsky and others cannot obviously be characterized by the rigid models available to the externalist. If natural languages are hostile to externalist treatments, then the meanings of natural language expressions serve as poor guides for ontological investigation, insofar as their meanings will fail to determine the referents of their constituents. This undermines the Realist's use of natural languages to settle disputes in metaphysics.

Keywords: Ontology, realism, polysemy, metaphysical methodology, externalism, internalism.

1. *Introduction*

Metaphysical investigation has, for the better part of the past century, been conducted by way of linguistic meaning. By tracing the meanings of expressions to their worldly extensions, metaphysicians aim to determine a sentence's ontological commitments by examining the purported worldly satisfiers of its truth-conditions. As a consequence, metaphysicians embrace—either explicitly or not—a particular view about linguistic meaning in order to render their investigatory practices coherent. My task here is to undermine this assumed view about

* My thanks to Erin Eaker, Steven Gross, J. Brendan Ritchie, and Alexander Williams for fruitful discussions about this project, and for comments on earlier drafts. I am also deeply indebted to Paul Pietroski for his repeated feedback on multiple drafts of this paper, and for the profound influence of his work on my own.

linguistic meaning, and thereby undermine the metaphysical methodology that assumes this troubled view about natural languages and their semantics.

Sider (2002, 2009, 2011) embraces the metaphysical view I hope to undermine, which holds that there is both an objective structure to reality, and a unique language that mirrors this structure:

...some candidate meanings [for an expression] ‘carve nature at the joints’ more than others, and it is part of the nature of reference and meaning that candidates that carve nature [closer to] its joints are more eligible to be meant. The meaning of a word, then, is the best candidate, where strength of candidacy is based on (1) fit with meaning-determining facts about the speaker or her linguistic community, and (2) intrinsic eligibility on the part of the candidate. (Sider 2002: xxi)

Sider embraces the first of these criteria in defense of using natural languages as metaphysical guides, holding that the use of natural language expressions constrains the space of any given expression’s possible “joint-carving” meanings. Applying such criteria makes sense for ontological investigation only if the relevant language has a semantics that determines the referents for expressions in that language. On this assumption, the quantificational commitments of true sentences in such a language constitute the domain of entities that stand to satisfy the truth-conditions of sentences in that language. And insofar as that language reflects the “jointedness” or structure of reality, this domain is just the domain of things that exist.

Such a language must have a semantics that is *externalist*. Put more precisely, this metaphysical methodology is committed to the following hypothesis pertaining to the (class of) language(s) it deems useful for metaphysical investigation:

(\mathcal{E}) For any expression e (in a given language L), the meaning of e determines e ’s truth-conditions.

For the externalist, the meaning of an expression determines its satisfaction-conditions or truth-conditions,¹ either by identifying meanings with truth-conditions, or by identifying the propositional content of a linguistic expression in a way that determines the truth-conditions of the expression. Likewise, the meaning of a word determines the mind-independent objects that constitute the word’s extension.²

¹ I use ‘truth-condition’ from here on, and in (\mathcal{E}) as short-hand for the much clumsier ‘truth-condition or satisfaction-condition’. The purpose for marking the distinction between these types of conditions is to mark the difference between sentences and sub-sentential expressions. Since sub-sentential expressions cannot be true nor false, they cannot have truth-conditions. But I trust that adopting this convention here will not lead to much confusion.

² The exact scope of this claim, and related externalist assumptions will be teased out with more precision in what follows, particularly in §2.1 and §4.1. To advertise a bit, the meaning of indexical expressions, even according to externalists, will fail to determine their truth-conditions in the absence of a (index providing) context. This sort of concern is (I think) not terribly vexing, in that externalism can accommodate

Despite the long-standing tradition in linguistics and philosophy of treating meanings as relations words bear to worldly objects, I will argue that natural language meanings are hostile to such externalist treatment. Most prominently, natural language expressions routinely exhibit a kind of *lexical flexibility* suggested by Chomsky and others, and this flexibility is ill-captured by the rigid models available to the externalist. Accommodating this flexibility under an externalist semantics either yields implausible ontological burdens on such theories, or belies good explanations for the relevant data, lending increased credibility to an internalist³ approach to linguistic meaning.⁴

Importantly, these findings bear on the projects of metaphysicians that see natural language meanings as suitable tools for ontological investigation. If natural languages are mind-dependent, and are poorly characterized by externalist semantic machinery, then the *Realist's* metaphysical methodology that reads ontology off of the qualificational commitments of natural language meanings is without foundation. The purpose of this paper is two-fold: first to undermine the basic externalist claim that the meaning of an expression determines its truth-conditions, and second, thereby undermine the Realist use of natural language meanings—and the intuitive judgments that rely on them—as useful tools in metaphysical inquiry. The goal is to undermine (\mathcal{E}), in the following Realist argument:

(\mathcal{E}) For any expression e (in a given language L), the meaning of e determines e 's truth-conditions.

If natural language meanings determine the truth-conditions and referents of their constituent expressions, then natural languages can play an important role in ontological investigation...

...since the meanings of true natural language expressions will pick out their real-world referents to populate the worldly domain.

If the meanings of terms in a natural language fail to determine their truth-conditions (and their constituents' referents), then investigating ontology by analyzing the meanings of true natural language expressions will yield indeterminate answers to ontological questions. The thought is that, with an adequately regimented natural language, one where vague and imprecise terms have been purged from the lexicon,

such cases. So while some considerations for the role of context in determining a sentence's truth-conditions are needed, the core Fregean ideal captured in (\mathcal{E}) is at the heart of externalism.

³ While the purpose of this paper is not to directly argue for such theories, see Pietroski (2008); Pietroski (2010); Pietroski (forthcoming) and Hinzen (2006); Hinzen (2007).

⁴ Some care will be taken regarding the terminology here, since 'meaning' has been used to identify numerous different properties associated with natural language expressions, utterances, interpretations, and the content of a bit of communication. In §4.1 these differences are spelled out, but for the time being, 'linguistic meaning' here is intended to identify the meaning properties of an expression that remain constant across various contexts in which that expression is used.

philosophers can make use of meaningful expressions in that language to do metaphysics, by tracing the meanings of terms in true expressions to their referents. It's this thought, and the externalist assumption (\mathcal{E}) presupposed by it, that I argue is troubled.

2. Setup

2.1. Taxonomy

The bulk of this paper is committed to advancing arguments against externalism. But before we proceed, some distinctions may be helpful. For purposes of taxonomy we should distinguish between internalist and externalist views about both *meanings*, and *languages*. As I understand the externalist's commitments, the *meanings* of natural language expressions are relations (of a particular sort). Meanings on internalist theories are non-relational, at least where one of the *relata* is a publicly available thing.⁵

The internalist/externalist distinction pertaining to *language* regards the ontological nature of languages. For the externalist view of language, languages are mind-independent things, while for the internalist, languages are aspects of the human mind. The logical space of externalist views regarding language and meaning are exhausted by adopting either 1) an internalist or externalist view about language, and 2) an internalist or externalist view about meanings.⁶ Thus one can be an internalist about language (IL) or an externalist about language (EL), holding that language is either in the mind or not.⁷ Similarly, one can be an internalist about meaning (IM) or an externalist about meaning (EM), holding that meanings either are relations between words and objects, or not. As such, the logical space of views is displayed in the following:

⁵ Drawing the distinction along relational lines under-determines the content of internalist theories of meaning, but for our purposes here, the under-determination is immaterial. Insofar as any externalist semantics relates publicly available things to natural language expressions (and internalist theories do not), if such a semantics cannot account for the linguistic data, this undermines (\mathcal{E}).

⁶ For the sake of completeness, one could also deny the existence of either languages or meanings. The motivations for either position aside, the Realist (the target of this work) would not welcome such a deflationary view.

⁷ This is an over-simplification. One can remain agnostic about the ontology of language, while denying an externalist conception (cf. Hinzen 2007: §1.5). One could also *deny* that there are languages at all, a claim many Chomskyans seem to endorse (cf. Chomsky 1986). Such theorists are considered internalists in the literature. I ignore this distinction for the purpose of simplicity, as nothing I say here trades on this distinction.

Table 1: Externalist and Internalist View

	Internalism Meaning (IM)	Externalism Meaning (EM)
Internalism Language (IL)	IL-IM	IL-EM
Externalism Language (EL)	EL-IM	EL-EM

2.2. Assumptions/Motivations

Anti-externalist arguments target both EL and EM theories. Though these two classes of arguments can be treated distinctly, they share a series of assumptions and motivations. Most notably they are inspired by a Chomskyan approach to language generally. In part this embodies a commitment to a naturalistic methodology. On this approach, language is treated as an object of scientific investigation, as a naturally occurring phenomenon, in principle no different than biological reproduction, combustion, planetary motion, or viscosity. This focus is not merely on empirical investigation, but that such investigation can make testable predictions, provide insightful explanations, and can be integrated with other scientific disciplines—most notably psychology and biology.

That humans have the ability to communicate the content of their thoughts *via* vocalization, and that we learn to do this in a short four years, are naturally occurring phenomena that beg for an (naturalistic) explanation. Explaining these phenomena in large part requires characterizing what it is to understand a language—put flat-footedly, one must “know English” in order to “use English.” As such, the study of language should seek to answer three questions:

- (i) What constitutes knowledge of language?
- (ii) How is knowledge of language acquired?
- (iii) How is knowledge of language put to use?

(Chomsky 1986: 3)

As Chomsky (1986) argues (and as we’ll see in §3), this commitment to naturalism and the guiding questions in (i)–(iii) rule out certain conceptions of language as viable candidates of study. They are excluded simply because language on these conceptions cannot be investigated through naturalistic means. Importantly, the veracity of these arguments depends on the success of the research program that insists on investigating language by naturalistic means. If the endeavors of such a research program bear no explanatory fruit, then that failure tells against treating language (and meaning) as a natural phenomenon. There is little doubt however whether the Chomskyan tradition in linguistics has failed in this regard (Baker 2002; Boeckx 2006; Piattelli-Palmarini et al. 2009).

2.3. Outline

The argument presented here is two-fold. First, I bring together various considerations, from different sources, which collectively offer sufficient reasons to doubt the truth of the claims adopted by externalists, in particular (\mathcal{E}). In the course of presenting these arguments I will both draw novel connections between the points offered by others, and consider (to then rebuff) externalist replies to them. The result then, is a series of considerations that collectively count against both the view that languages are mind-independent objects (i.e. EL theories), and the externalist thesis (\mathcal{E}) as it pertains to natural languages (i.e. EM theories).

The arguments offered here do *not* show that externalism is false. The purpose of gathering together this evidence is to show that the externalist *hypothesis* (\mathcal{E}) is contentious and troubled, and not an obvious truth to be taken as the starting point for other domains of inquiry.⁸ The second part of the argument here presses this point, since much metaphysical investigation *assumes* the truth of this externalist thesis. If the considerations offered here against the externalist proposal are indeed as troubling as I claim, then the Realist that conducts ontological inquiry by way of natural language meanings is burdened with those troubles.

I begin by addressing externalist views that treat languages as mind-external objects (i.e., EL theories), by arguing that such views forestall attempts to address central questions about the human capacity to use and acquire natural languages. In §4 I turn to externalist views about meaning (i.e., EM theories). Because of the long standing externalist tradition in linguistics and philosophy of language, I begin that section with a lengthy clarification, indicating the sort of linguistic phenomena that I do *not* think are troubling for the externalist. So clarified, I argue that the *lexical flexibility* exhibited by broad swaths of natural language expressions are poorly explained by the externalist hypothesis (\mathcal{E}). I consider possible externalist responses to the challenge presented by such flexibility before turning to the assumed role of mental content at play in EM theories. There I stress the complex relationship between linguistic expressions and the concepts they purportedly express, noting that the commonly held *labeling* theory of linguistic meaning—whereby words are merely labels for concepts—faces two profound difficulties. After presenting this rather large body of evidence against various externalist commitments, I conclude by arguing that these considerations against externalism present a fundamental challenge for Realist metaphysical methodologies, by showing that they assume the externalist theses this evidence renders contentious.

⁸ It's worth remembering here the oft-quoted claim from Lewis (1970), the progenitor of contemporary metaphysics, whereby he boldly asserts the externalist hunch, with little argument: "Semantics with no treatment of truth-conditions is not semantics."

The consequence for the Realist is that her use of natural language meanings to address metaphysical questions is unjustified.

3. *Languages as Objects*

3.1. *Arguments Against EL Theories*

There are two conceptions of the ontology of languages, an externalist (EL) view of language and an internalist (IL) view of language. The former view construes a language as a mind independent (abstract) object. Languages on this conception are abstract structures relating mind-independent objects to terms, words, or expressions—artifacts in some sense, that we use to denote (other) objects. Accordingly, one understands a language when they can identify and use the abstract structure that most sensibly coheres with the usage in their linguistic community. This conception of language can be found in Lewis (1970, 1975):

What is a language? Something that assigns meanings to certain strings of types of sounds or of marks. It could therefore be a function, a set of ordered pairs of strings and meanings. (Lewis 1975)

A commitment to naturalism speaks against thinking of languages as the abstract objects described here. In treating a language as a mind external object, as Lewis (1975) puts it, “in complete abstraction from human affairs” one wants to know how humans can come to understand or “know” languages so construed (p. 19). That is, on this Lewisian characterization, languages are abstracta: functions that take us from symbols to truth values, combined with a grammar that delineates how these symbols can be combined in acceptable ways to form interpretable expressions (or strings). If a naturalistic approach to language seeks to answer (i) what knowledge of a language amounts to, a Lewisian treatment of language renders this question intractable. As an infinitely-membered set of ordered pairs of expressions and their functional meanings, this conception of language not only gives us little direction as to how to answer (i), but seemingly gives too sparse a collection of resources to answer the question at all. Put more tangibly, all children (placed within a linguistic environment) have adult-like competence with a natural language by the age of four. A child when placed in a community of English speakers will come to “know English” by the age of four. On a Lewisian view, for a child to “know English” they must first decide which set amongst an infinite array of (infinitely-membered) sets of expression-function pairs is the English set, and then second, they must bear the right kind of epistemic relation to that set. As such, to explain what knowledge of a language (so construed) amounts to requires an account of the sort of relation that a human can bear to an abstract entity such that this is the sort of relation a four-year-old child can enter.

The worry here is not that, given certain metaphysical commitments to nominalism, any theory committing us to *abstracta* is off base. The worry Chomsky presses pertains to the conditions for explanation, and particularly whether certain conceptions of language (whatever their metaphysical commitments) forestall viable strategies to answering fundamental questions. Treating languages as abstract entities is problematic not because they are abstracta, but because abstracta *qua* objects of knowledge bear mysterious epistemic relations to human minds. As such, to explain what knowledge of a language (so construed) amounts to requires an account of the sort of relation that a human can bear to an abstract entity, and in particular a set of ordered pairs, such that this is the sort of relation a four-year-old child can enter. How one proceeds to answer these questions seems hopelessly unclear.⁹

Chomsky puts a related point about language acquisition in terms of “legibility conditions” on a natural language. If a child is to come to have “knowledge of a language,” they must come to represent that language in their mind/brain. For a given child to have “knowledge of English,” they must have come to represent both the grammar of English—the algorithms by which one can combine lexical items to form larger expressions (or sentences)—and at the very least, some internalized list of those lexical items that combine syntactically. In whatever way this information is encoded in the mind/brain, it must fit into the architecture of the mind/brain. If our four-year-old has figured out, or “knows”, which grammar (as an abstract object on an EL theory) is the English grammar, she has represented such a grammar in her mind/brain. When she wants to utter an English sentence this grammatical knowledge must be applied (to the lexical items she also “knows”) in such a way so that her articulatory systems can make the right sort of audible noise. That is, the representations she builds using her grammatical knowledge must encode information in a way that her articulatory system can make sense of—those instructions must be *legible*. Thus the structure of the abstract object that is English must abide by such legibility conditions, coming not only from the articulatory system, but from any other aspect of the mind/brain that the child’s linguistic knowledge must interact with. But once we recognize that the structure of the abstract object *English* is beholden to the legibility

⁹ A Lewisian might argue that such an explanation need not be directly forthcoming for the project of building an externalist semantics for externally construed languages to proceed. After all, humans somehow manage to learn mathematics, and the best account of the ontological nature of mathematical language is decidedly abstract, and set-theoretical. So clearly (the thought goes) humans can stand in the relevant epistemic relation to abstract objects, and in particular abstractly construed languages. Notice however, that four-year-old children do not exhibit mastery and competence with the language of mathematics, even if they do have innate mental structures that aid them in acquiring proficiency with numerosity (Carey 2009: Ch. 4). This contrast between children’s felicity with mathematical and natural languages requires an explanation, and that explanation is precisely what the Chomskyan challenge to EL views demands.

conditions imposed by the human mind/brain, this defeases the motivations for thinking of a language in this abstract way Chomsky (2000).¹⁰ Taken together, considerations of acquisition and legibility pose serious, and as far as I can see, unanswered challenges for EL theories.

None of this shows that these *abstracta* do not exist. There could well be, in addition to the mental structures hypothesized (and studied) by the Chomskyan tradition, functions filling an infinite Fregean hierarchy mapping objects of one kind (say <*e*>) to objects of another kind (say <*l*>). But these functions must earn their keep. We are told that we should believe in these *abstracta* because they are essential to successful explanations of linguistic phenomena (Lewis 1986: Ch. 1). The point here about legibility conditions is that these structures serve no explanatory purpose, and indeed present explanatory obstacles, in explaining how a child comes to “know” a language. Thus, the externalist owes us some indication of what these objects are meant to explain.

One possible explanatory virtue of adopting an externalist view about languages (and meanings) is that it yields an intuitively plausible account of successful communication. If languages (or their meanings) are external entities, then successful human communication is explained by the mutual relatedness of individuals to the language(s) they know. That is, two humans can succeed in communicating because they are related to, or come to understand, the very same *thing*—a language. But this explanatory virtue depends on the availability of a cogent specification of the individuation conditions for the external languages that individual speakers come to grasp. The prospects for these conditions, I contend, are grim, at least for the Realist.

3.2. *Grim Prospects for Linguistic Objects*

Commitments to naturalism motivate the ontological arguments against EL theories above. EL theories hold that natural languages are objects-in-the-world, whose existence is independent of minds. As such, expressions of a natural language are likewise objects-in-the-world. As we saw in the previous section, naturalist inquiry into the acquisition of natural language compels the externalist to defend the utility of EL conceptions of language. An externalist might reply, indicating that treating languages as publicly available objects yields a plausible account of linguistic communication. If languages are mind-external

¹⁰ Chomsky also argues that treating languages as abstract structures, akin to the formal languages of mathematics, renders aspects of natural languages inexplicable, in principle. Chomsky uses two examples, ‘imperfections’ in natural languages, to highlight the mismatch between the structure of natural languages and formal languages: 1) that natural languages have uninterpretable features, and 2) the displacement property (Chomsky 2000). Even if treating natural languages like formal ones leaves room for explaining such features, the point here is that there is no good justification for stipulating at the outset of investigation that the object to be investigated must meet the (optimality) conditions of a formal language (especially if even superficial differences speak against such stipulation).

things, communication between speakers can be understood by way of their respective relations to the same publicly available thing—the language they all are said to know. One strain of internalist arguments presses this ontological claim, showing that there is no scientifically respectable notion of ‘natural language expression’. More plainly, these arguments contend that mind-independent words (and thereby, mind-independent languages) do not exist, and thus cannot stand in any (reference/meaning) relation to any object in the world.

One motivation for thinking of languages as external is to account for communicative success. In order for two speakers to understand one another, they must know the same language, and this (so goes the argument) can be explained if there is some single mind-external thing they both know. Knowledge of language on this view is had when a speaker bears the right sort of relation to this external object. Thus two speakers that “know English” can make sense of each other’s speech because they bear the same relation to the same mind-external object. Thus when you use a sentence(-object) of English, I understand this sentence because I am related to that *same* English object in the *same* way you are, *via* knowing. However, this requires that languages can be differentiated ontologically, insofar as you and I must bear the same relation to the *same language*.¹¹

For purposes of differentiation, “common sense” methods will not do. As Chomsky notes, common sense treats Dutch and German as distinct languages, despite the fact that people “who live near the Dutch border can communicate quite well with those living on the German side...” (Chomsky 2000: 48). If treating languages as mind-external objects is meant to explain communicative success, then the fact that speakers of “different languages” can communicate linguistically¹² is unexplained by such theories. The common sense notion of language gets the extension wrong in the other direction as well, insofar as the Mandarin and Cantonese spoken “dialects” of “the Chinese language” are mutually

¹¹ There is a fundamental problem with this conception of language, namely that any particular human will fail to *completely* know any such “language.” Externally construed as an infinite set of expression-meaning pairs, no individual will come to have full knowledge of a language, at least if such knowledge is construed as knowing what the members of this set are. Given the limits on human cognition, no individual could know this infinitely long list of pairs. At best, we must have an incomplete knowledge of such a “language.” But the account of communicative success above assumes that two speakers stand in the same relation to the same object. If a language is an infinitely large set of sentence-meaning pairs, two speakers of that language either: (a) stand in a different (incomplete) knowing relation to that same set, or (b) stand in the knowing relation to different subsets of that set. Thus if appeals to some external entity are meant to aid in explaining how humans manage to successfully communicate, such an explanation will have to succeed despite the fact no human can fully know such a language. See Dummett (1978, 1993) for discussion.

¹² Humans, and other animals, communicate in many non-linguistic ways: body language, facial expression, gesture, etc. The important cases for this point are those instances of communication that are clearly effected with language.

unintelligible. The EL-view here fails to explain why these speakers of “Chinese” cannot communicate *via* spoken language, since (according to common sense) the relevant populations both “know Chinese.”

Since the common sense division of languages will not serve the purpose of explaining communicative success, another means of distinguishing languages (as objects) is in order. One might appeal to the *elements* of languages to distinguish them. On this proposal, two languages, (say) English and French, are distinct because of the differences between the elements that constitute them—one contains words like ‘photographer’, ‘apartment’, and ‘cat’ while the other contains words like ‘photographe’, ‘appartement’, and ‘chat’. This move requires that these elements differ along some important dimension, such that the first three belong to English and the last three belong to French. Notice that appealing to usage will not be helpful. Defining ‘French’ and ‘English’ by indicating that speakers of French use the latter and speakers of English use the former is viciously circular.

An EL theorist must make use of some other property that these words share that marks the boundary between English and French. But to what properties could an EL theorist appeal? The sonic properties of these words seem like bad candidates. ‘Cat’ and ‘Chat’ have more in common in this regard than ‘photographer’ and ‘cat’ do. Further, considering the variation seen in pronunciation across speakers of the “same word”, appeal to such properties will not distinguish words as to cohere with the communicative motivations for EL theories:

To take an example, why are ‘fotoGRAFer’ (said in Bombay) and ‘foTAH-grafer’ (said in Toronto) the same word, yet ‘foTOgrafo’ (said in Buenos Aires) is not the same word as the former two? (Stainton 2006: 918–919)

To explain communicative success and failure, the EL theorist posits the mind-external object *English* that our Bombay speaker and Toronto speaker both “know”, distinguished from other languages based on the elements of that language—words like ‘photographer’. In this example, the EL theorists wants to say that we have two words here (as opposed to one or three), one in English and one in Spanish. Appealing to the sonic properties of (utterances of) words here will clearly not help such a theorist, given variations in pronunciation.¹³ The point is familiar to phonologists, namely that there is no *sui generis* cluster of sonic properties that utterances of a word share in common. What ‘fotoGRAFer’ and ‘foTAHgrafer’ share in common is the manner in which they are represented by humans, which involves features of the system for creating phonological representations from environmental noise.¹⁴ There are then no word-objects that can be differentiated without appeal to structures of the human mind/brain—or no such objects that can be investigated by naturalistic means.

¹³ Appeals to orthography will not be any more helpful here given that illiterate individuals can communicate using spoken language quite well.

¹⁴ See Bromberger & Halle (1995) for discussion.

These problems are particularly trenchant for theorists committed to a Realist position. For the Realist the naturalist perspective is not negotiable. If there is no mind-external, naturalistically respectable notion of ‘natural language expression’, and thus no worldly objects of that sort, then there can be no (semantic) relations between words and objects. Such a Realist cannot fall back on a kind of pluralism or fictionalism about words, because their methodology prohibits such a retreat.

Thus a commitment to naturalism will preclude certain notions of language, specifically those that construe natural languages as mind-independent abstract entities. However, this naturalistic commitment does not preclude a semantics for an I(L)-language from having an externalist character. Even if language is properly construed as an aspect of the human mind, the expressions of that language, seen as mental representations, can still have contents that are cashed out in terms of mind-independent objects. Such an account of meaning is still importantly externalist. This view pushes the externalist to adopt two positions: first, an EM conception of linguistic meaning, and second an externalist account of mental content. In the remaining sections I discuss problems for both of these views. I turn next to the various arguments put forward for why (natural) I(L)-language meanings are hostile to externalist treatment.

4. Compositional Referential Semantics and Natural Language

A theory of meaning for a natural language should provide a means for pairing sentences of a natural language with their meanings. Various semantic theories accomplish this goal in varying ways, but externalist semantic theories insist that such a pairing requires relating sentences to objects in a (worldly) domain. These various EM theories adopt the thesis in (E). But a theory of meaning must also explain the distribution of competent speaker judgments about natural language constructions. This includes (but is not limited to) judgments about inferences speakers draw between expressions, and the meanings they are (not) apt to assign to expressions. A theory of meaning ought to offer a plausible explanation for this data.

The primary strain of anti-externalist arguments I will press illustrate that an externalist semantics cannot obviously accommodate the relevant data, or that in accommodating the relevant linguistic data such a semantics must take on board independently implausible assumptions. As a means of articulating the data points I contend EM theories fail to capture, I’ll begin this section by briefly indicating what these criticisms are *not* arguing.

4.1. *How Not to Argue against Externalism*

There is a line of criticisms attributable to Austin (1962), Strawson (1950), and Wittgenstein (1953, 1972) regarding externalist semantics summarized by indicating that words do not refer/denote, *users* of words do.¹⁵ One way of explaining this point is to highlight the role of indexicals (and demonstratives) like ‘I’, ‘you’, ‘here’, ‘now’, and ‘that’ in determining the meanings of expressions that contain them. The meaning of indexicals are intimately tied to the context of their use. Even if we accept that sentences like (1) have constituents with referential meanings,¹⁶ sentences like (2)–(4) do not, because they are importantly incomplete (nonsensical for Wittgenstein¹⁷) outside of their use:

- (1) Kiruna is foggy.
- (2) I think that should go there.
- (3) I am here.
- (4) I am a philosopher.

Sentences with indexicals, like (2), (3) and (4), have different meanings when uttered by different agents (at different times, different places, and using different gestures). In such cases, context plays the role of determining the referent of the indexical constituent(s) of these expressions. Call this the role played by *semantic context*. The meaning of these expressions determine their truth-conditions, in a context that supplies coordinates for the indexes proposed to be part of their interpretations, and thereby maintains the externalist idea that meanings determine the truth-conditions of expressions (Kaplan, 1977 [1989]). Such cases do *not* undermine EM theories.¹⁸

¹⁵ To quote Strawson (1950): “Referring is not something an expression does; it is something that someone can use an expression to do.”

¹⁶ Though this is far from obvious, despite what traditional semantics textbooks might suggest. Even if the conditions that satisfy the predicate ‘is foggy’ are codified, consider the ontological status of a city that moves from one location to another location two miles to the east (Rolander 2013). And even if those metaphysical difficulties can be met, it is far from clear how such a view can account for sentences like “The tallest mountain in Sweden, Kebnekaise is situated 100 km or 62 miles from Kiruna. ...Nowadays the town [of Kiruna] is not relying solely on the mine” (girontravel.se, 2013). While a city construed as a spatio-temporal object might well be situated some distance from some other object, and might even survive relocation, such an object is not of the sort that relies on anything, much less revenue. See Chomsky (2000) for examples of this sort regarding London.

¹⁷ See Wittgenstein (1972: §10).

¹⁸ Strawson and Wittgenstein would surely disagree. Broadly, the point pressed by many ordinary language philosophers is that the sort of distinction presented here between semantic and cognitive context, and indeed the semantics/pragmatics distinction generally, is misguided. My brief treatment of these concerns here is meant merely to clarify the arguments proffered in the remainder of this section. Even if the externalist can address the problems raised by Strawson and Wittgenstein, possibly in the manner indicated here, many trenchant problems remain. My thanks to an anonymous reviewer for raising this concern.

There are other ways in which the context of an expression's use can impact or "change" the meaning of the expression. Consider the following expression:

(5) Barack Obama is human.

The "literal" meaning of this expression, on an externalist understanding, indicates that some particular individual has a particular property. I can use this expression to convey a thought, the meaning of which is that some individual is a *homo sapiens*, on those occasions where I intend to convey the *linguistic meaning* of the expression. The related Austin-inspired point regarding the (externalist) meanings of expressions contends that because we use language in non-literal ways, and further that such non-literal usage is pervasive, the meanings of expressions are the conditions of their use. Thus what a sentence *means* follows from the contexts in which a speaker can felicitously use the sentence. For example, one could use (5) in various ways:

(5') Barack Obama is human. *Context*: Obama has made some mistake.

In uttering (5) in the context of Obama's mistake in (5'), I do not merely intend to communicate information about Obama's place on the phylogenetic tree. My usage communicates (or intends to communicate) the linguistic meaning acontextually attributable to the expression in (6):

(6) One should refrain from feelings of disapprobation toward Obama in this instance in light of the fallibility of humans.

Call the role context plays in such cases *cognitive context*.¹⁹ This non-literal usage of language is pervasive.²⁰

However, this feature of language does little to undermine the externalist program in semantics. That we can use sentences to convey thoughts that do not match the intuitive literal meanings of expres-

¹⁹ For a discussion of this differing role of context, and the distinction between semantic and cognitive context see Bach (1999, 2004).

²⁰ This feature of natural language seems to be at the heart of Strawson (1950), which is often misunderstood as merely indicating the context sensitivity of indexical (and demonstrative) expressions. While one way of making "different use of the same sentence" is to use an indexical expression in different contexts, this is but an instance of a more general phenomenon (Strawson 1950). The general point for Strawson is that the proposition expressed by a speaker, and thus whether what is spoken is true, depends quite heavily on the context of utterance (and the use of the sentence)—a point Russell seems to miss. Strawson focuses on the context sensitivity of indexicals, largely embedded in definite descriptions, because he argues that Russell's theory of descriptions, which sacrifices the connection between grammatical and logical form for the sake of sentential truth preservation, does so needlessly (Strawson 1950). The theory is not needed, according to Strawson, because *sentences* are not true or false, *uses* of sentences in utterances are. Russell's reply seems to miss this substantive point: "As regards 'the present King of France', [Strawson] fastens upon the egocentric word 'present' and does not seem able to grasp that, if for the word 'present' I had substituted the words 'in 1905', the whole of his argument would have collapsed" (Russell 1957).

sions *requires* that expressions have static linguistic meanings that are context-independent.²¹ And this literal meaning could be externalist, even if the communicative intention of a speaker using such a meaning heavily depends on cognitive context.

The gap between linguistic meaning and the thought inferred by an audience to a speech act (or utterance) does not present a problem for truth-conditional semantics. However, some utterances closer to the semantic-pragmatic boundary seem more troubling. Consider the following examples:

- (7) John is too smart [for this job].
- (8) John finished [writing/playing] the sonata.
- (9) John is ready [for class/to go home/...].

Each of these sentences, though acceptable to competent speakers of English, seem (in some sense) incomplete, as indicated by the bracketed content. For the externalist, the linguistic meaning of expressions of the form in (9) [NP-COP-XP] predicate properties to the individual(s) denoted by the noun phrase. ‘John is awake’, for example, predicates of the individual denoted by ‘John’ that he is not asleep. But the meaning of (9) does not predicate some general property of *readiness* to *John*.²²

Interestingly, these sentences and their completed counterparts stand in a relationship that exhibits features typically attributed to semantic properties, yet seem to be driven by pragmatic inferences. The

²¹ In uttering a sentence, I intend to communicate some thought or other. The task of my audience is to infer this thought from the sentence I used in my act of uttering. Given my overt intention to communicate a thought, my audience must identify the intended thought, in some way or other (Grice, 1957 [1989]). Contextual cues, shared biological sense modalities, a common presupposed set of knowledge, and other aspects of the uttering act all constitute the evidence available to my audience in making the correct inference about my intention. When communication is successful, they ascertain my actual intention. But paramount among the evidence considered in this inferential move is the choice of sentence used in the utterance. That I utter (5) and not (say) its negation in the context indicated in (5’) seems to matter, and it matters precisely because the sentence has a static linguistic meaning. Were the sentence void of any literal meaning, it could play no role in my audience’s inference making. That I cannot communicate the thought in (6) by using *any* expression I choose illustrates this point. There are limits on what a speaker can reasonably expect his audience to infer about his communicative intention on an occasion of utterance, and largely this is because expressions have static linguistic meanings.

²² Though this seems to be the view of Cappelen & Lepore (2005). This view however, fails to explain what needs explanation in these cases. Namely, competent speakers of English treat (9a) and (9b) as having the same meaning. The disquotational account of meaning fails to capture this data in any non-stipulative way. In fact, for Cappelen & Lepore (2005) the proposition expressed by a sentence is the disquoted sentence once we “disambiguate every ambiguous/polysemous expression in [the sentence]” (p. 145). However, the presumption that cases of ambiguity and polysemy can be resolved, *prior to* giving a semantics for the expression either denies that the data need explanation or denies that there are any data there to explain. For (9a) and (9b), that speakers treat these sentences as having the same meaning is ill-captured by a theory that insists they express distinct (disquoted) propositions.

relationship between (9a) and (9b) is different than the relationship between (5) and (6).

- (9) a. John is ready.
- b. John is ready for it.
- (5) Barack Obama is human.
- (6) One should refrain from feelings of disapprobation toward Obama in this instance in light of the fallibility of humans.

Cases like (9a) and (9b) present a problem because the role that context plays seems to be more general (or cognitive), while exhibiting (semantic) entailment patterns that pragmatic cases do not. To complete (9a) and arrive at (9b), my audience has to make use of knowledge not provided by the semantic context, specifically knowledge about *readiness*. Yet (9a) and (9b) exhibit a mutually entailing relationship that pragmatically inferred thoughts rarely have. Regardless of how (semantic) context provides the expression ‘it’ with the relevant event that John is ready for in (9b), this determination will (for the externalist) determine the truth-conditions for the expression. But any context in which the provided event makes (9b) true, also makes (9a) true.

The worry then is this: (9a) and (9b) seem to have the same truth-conditions, yet appear to have different meanings, insofar as (9a) is incomplete and (9b) is not. For the externalist, the truth-conditions of an expression are determined by its meaning. That they come apart in these cases is troubling for the externalist view.²³

The anti-externalist arguments in the remainder of this section are not the (now) traditional worry associated with Strawson, Austin, and Wittgenstein that a single expression can be used in a variety of ways to express a variety of thoughts. Nor is the worry expressed by internalists that expressions with indexical constituents require context to determine their truth-conditions. Such points do not speak against an externalist semantics (though they indicate that (\mathcal{E}) requires some clarification regarding the determination relation). The point that internalist worries stress is that, while the meanings of indexical expressions might be well-captured by appeals to (something like) the content-character distinction, they are a special case of a much more general phenomenon that is misrepresented by such treatment, and exhibited by case like (7)–(9). Natural language expressions exhibit a kind of *lexical flexibility* that is not isolated to a few problem cases to be addressed by intricate logics, but is a ubiquitous feature of natural languages—one that is importantly misrepresented by Kaplanian treatment. It is to this phenomenon that I now turn.

²³ Stanley (2000), for example, notes this trouble and offers a solution to cases of this kind by introducing unarticulated syntactic constituents. My purpose here is not to evaluate the merits of every externalist reply to such cases. I introduce them here merely to suggest the sort of worries I will (not) focus on in the remainder of this section.

4.2. *Lexical Flexibility*

An externalist semantics, or EM theory, must be capable of explaining or accommodating the distribution of meaning assignments competent speakers of a language give to expressions. In the previous section we reviewed proposed counterexamples to (E). In response to these cases (E) could be maintained, by adding characters to the meanings of expressions, as is done for indexicals. In this section, we will consider phenomena that cannot be so easily accommodated, namely those related to *lexical flexibility*.

Chomsky (1977) marks a distinction between various ways in which the meaning of an expression can be multifarious, distinguishing between expressions that exhibit *flexibility*²⁴ from those that exhibit *ambiguity*. The English word ‘trunk’ is ambiguous, with meanings used to denote both luggage, and a part of an elephant (not to mention tree parts, and humans parts). This kind of multifarious meaning is importantly different than the kind exhibited by expressions like ‘book’. Compare the following expressions:

- (10) John wrote a book.
 (11) This book weighs five pounds.

The use of ‘book’ in (11) is used to denote a particular, concrete (heavy) book. For (11) to be true there must be a contextually relevant physical thing with a particular heft. However, for (10) to be true there is no such requirement. As Chomsky notes, John could have the book composed in his mind, having never deployed pen and paper. This might lead us to conclude that the multifarious meaning of ‘book’ is like the *ambiguity* of ‘trunk’: we have two distinct (homophonous) lexical items, ‘book_a’ corresponding to the *abstract* usage in (10) and ‘book_c’ corresponding to the *concrete* usage in (11).

However, if this treatment of ‘book’ is apt, we should expect uses of ‘book’ to behave like uses of ‘trunk’, since ‘trunk’ *is* ambiguous is the way ‘book’ is purported to be on this explanation. That ‘book’ is less

²⁴ Often this phenomenon is termed ‘polysemy’ in the literature. I avoid the term here for two reasons, one priggish, and one substantive. The priggish reason is that the term ‘polysemy’ indicates (by its roots) that a polysemous word has multiple meanings. This is true for such terms of course, but ambiguous words are also ones with multiple meanings. Thus the contrast in the literature between polysemous terms, which are troublesome for mainstream semanticists, and ambiguous terms, which are not troublesome, is misrepresented by this use of terminology. The substantive reason for my usage here is that lexical flexibility is a property that applies rather broadly, and manifests with different semantic behavior in different contexts, two of which I discuss here. However, the explanation for the various manifestations of polysemy can be unified by internalist proposals (Pietroski 2005: §3.2). See also Pietroski (forthcoming). Lastly, my usage does not ignore the role of pragmatics in understanding the import of the examples presented below, but rather insists that the phenomena to be explained are semantic (as suggested by the discussion in §4.1).

well-behaved suggests that the semantic relationship between uses of ‘book’ is of a different sort:

- (12)a. This book, which John wrote, is five pounds.
 b. John wrote a book, this is it, and this book is five pounds.
 c. John wrote a book_a, this is it, and this book_c weighs five pounds.
- (13)a. This trunk, which Jumbo grew, is full of clothes.
 b. Jumbo grew a trunk, this is it, and this trunk is full of clothes.
 c. Jumbo grew a trunk_a, this is it, and this trunk_i is full of clothes.

The expression in (12a) is (roughly) synonymous with (12b), using ‘book’ in the two distinct ways discussed, as reflected in (12c). But (13a) is not synonymous with (13b), at least not if we interpret (13b) as making use of both lexical expressions of ‘trunk’ as in (13c). They are synonymous if we imagine Jumbo’s nose full of textiles, but this interpretation is not available for (13c). Importantly, this difference is not attributable to the syntax of these various phrases, as (12a) and (13a) appear in the same syntactic frames.

That these two expressions (‘book’ and ‘trunk’) behave in semantically disparate ways in relative clause constructions counts against a semantics that treats them as formally similar. That is, this behavior suggests that we do *not* treat the relationship between uses of ‘book’ as ambiguity (i.e., homophony). As Chomsky concludes:

Thus [in cases like ‘book’] we have a single formal element with a fixed range of meaning, and relativization is possible, despite the shift of sense. But in the case of ...[‘trunk’] (idiosyncratic ambiguity) we have two formal elements ...with the same phonetic form. (Chomsky 1977)

In cases like ‘book’, the lexical entry has a range of interpretations, exhibiting a *flexibility* that permits a kind of mixed use as in (12a). In cases like ‘trunk’ there are two lexical entries that are *homophonous*, each with distinct and unrelated meanings, rendering mixed use interpretations of expressions like (13a) unavailable. The trenchant point is that the flexibility exhibited by ‘book’ is pervasive in natural languages, and poorly captured by theories that treat them as cases of homophony (as seemingly EM theories must).²⁵

There are two kinds of flexibility explored in the remainder of this section, one based on non-linguistic knowledge, and one based on ontological type. The truth-conditions for some natural language expressions are not determined by the referents of their constituents, and the manner of their composition. To determine their truth-conditions, competent language users must deploy non-linguistic knowledge of a general sort—too general to be considered the semantic context of utterance. The second sort of flexibility permits expressions that ap-

²⁵ As a limiting case of this strategy, which treats color terms as massively homophonous at the granularity of *use*, see Rothschild & Segal (2009).

ply multiple predicates of different types to a single noun phrase. The would-be externalist meanings of such expressions require objects of a(n) (impossibly) bizarre sort.

4.2.1. *Non-linguistic Knowledge*

Consider the following expressions:

(14) Football games are played by jerks.

(15) Residential houses are robbed by jerks.

Sentences like (14) and (15) highlight the fact that we bring extra-linguistic knowledge to bear on linguistic expressions in determining the truth-conditions for sentences. The information contained in the *linguistic* properties of expressions like these (even when combined with their semantic contexts) do not determine the truth-conditions of those expressions. As such, meanings cannot be (or cannot determine) truth-conditions, and thereby cannot be externalist.

On any EM theory the difference in the truth-conditions of any two sentences (*modulo* semantic context) must be a consequence of the difference in either their constituents or the manner in which those constituents are syntactically related. So, given that (14) and (15) appear in the same syntactic frames, and that they differ only with regard to the two constituents ‘football-games’/‘playing’ and ‘residential-houses’/‘robbed’ respectively, whatever (14) indicates is true of the relationship between *football-games*, *playing*, and *jerks* should, according to (15), hold true for *residential-houses*, *robbing*, and *jerks*.²⁶

However, the expression in (14) means that *every*²⁷ football game is played by jerks, while the expression in (15) emphatically *does not* mean that *every* house is robbed by jerks. The information essential for deriving the truth-conditions for (14) involves the tight relationship between *games* and *playing*, namely that there can be no unplayed game—a relationship that does not hold between *houses* and *robbing*. But this information is not a linguistic property of the expression, and not part of the linguistic meaning of the expression. Thus in deriving the truth-conditions for (14) a competent speaker of English relies on information not present in the expression (nor even the semantic context). If meanings are (or determine) truth-conditions, then the truth-conditions of any meaningful sentence will be determined by the expression alone. But such a theory will fail to explain the differences in meaning between (14) and (15) not captured by the difference in their constituency.

²⁶ See Chomsky (1975) for similar examples.

²⁷ Admittedly, these expressions are generics, and speaker judgments in this domain are (seemingly) not concordant. However, even if the expression in (14) is not interpreted with a universal quantifier, (14) indicates (at a minimum) that *most* football games are played by unsavory individuals. In contrast, (nearly) no one will interpret (15) as indicating that most residential homes are robbed, even if most of those robberies are perpetrated by jerks. Since this difference is not syntactic, the externalist is burdened to explain why the relationship between the VPs and the NPs in these two expressions is different.

This general phenomenon is not limited to generic expressions. We bring non-linguistic knowledge to bear on linguistic expressions in other ways that do not seem to rise to the level of pragmatic inferences, but are also not a function of semantic context. The differing contribution of expressions like ‘coffee’ as a predicate in ‘coffee drink’ and ‘coffee grinder’ present a *prima facie* problem for the externalist. If the semantic contribution of an expression can be recovered by the semantic contribution of its constituents (plus the means of their composition) the meaning of ‘coffee’ in these expressions should make the same contribution across uses. But a coffee drink is one that is *composed of* coffee, while a coffee grinder is not made up of coffee at all. While these are facts that *any* theory of meaning needs to explain, the externalist will have particular difficulty dealing with this problem, insofar as the contribution of ‘coffee’ on such a theory is exhausted by its reference to a property—and in particular, a property instantiated by the indicated objects in the various expressions in which it functions as a predicate. But there seems to be no obvious single candidate for the needed property in this coffee-case. The best candidate properties for the would-be denotations of ‘coffee’ in these two expressions seem profoundly different: *being-composed-of-coffee* and *used-in-the-production-of-coffee*. The difficult task for the externalist is not only in pinpointing the relevant *single* property in such cases, but also in constructing an account of how speakers come to triangulate on such (non-obvious) properties in the many cases that exhibit these features (e.g. ‘metal shears’, ‘home loan[/inspection]’, ‘rain delay[/coat]’, ‘blue marker’, etc.). This phenomenon is pervasive in natural language, and not easily explained by the externalist.

4.2.2. *Ontology and Satisfaction*

The lexical flexibility exhibited by natural language expressions cannot be accounted for by an EM theory that demands that the meanings of expressions determine a unique referent, as a mind-independent object. Consider the following two English expressions

(16) The Hirshhorn-Museum is bankrupt.

(17) The Hirshhorn-Museum is a cylinder.

A competent speaker of English could think that these expressions are true. An EM theory accounts for this fact by indicating that such a speaker takes the following conditions to hold in the world: 1) there is an object, the one ‘Hirshhorn-Museum’ denotes, which satisfies the conditions for ‘is bankrupt’; and 2) there is an object, the one ‘Hirshhorn-Museum’ denotes, which satisfies the conditions for ‘is a cylinder’. This would require that there is some single object, denoted by ‘Hirshhorn-Museum’, that can be both bankrupt and a cylinder.

The predicate in (16) requires that this object be a financial institution, understood through an array of socio-economic notions. Whatever

these notions demand about the nature of financial institutions, be they collections of individuals or something more abstract, such objects do not seem to have a *shape*. Similarly, the kind of object that would satisfy the predicate in (17), in this case a building, is seemingly not the sort of thing that can have financial troubles. Building projects can have fiscal crises, but buildings seemingly cannot. The EM theorist then either owes us an account of the kind of object that can satisfy the predicates in both (16) and (17), or they must explain why this problem does not generate in the first place.

Taking the latter strategy, an EM theorist might appeal to the presence of ambiguity in natural languages, as in the following:

(18) The geese are by the bank.

The expression in (18) is ambiguous. Much like we saw with ‘trunk’ earlier, we can treat ‘bank’ as homophonous, indicating the distinct lexical items ‘bank_f’ and ‘bank_b’. The ambiguity of (18) is then explained by appeal to the homophony of these two distinct lexical items. The expression in (18) simply fails to determine which ‘bank’ is being used, and as such the expression can have different meanings based on which item is intended.

Likewise, one could argue that we really have two lexical entries for ‘Hirshhorn-Museum’, one that denotes the institution, and another that denotes a building. We can represent this difference between (16) and (17) as

(16’) The Hirshhorn-Museum_f is bankrupt.

(17’) The Hirshhorn-Museum_b is a cylinder.

Since ‘Hirshhorn-Museum’ identifies two distinct (though homophonous) lexical items, contextual information determines which item is used in (16) and (17) respectively, preserving the distinct meanings of the expressions, while assuaging the worry that an EM semantics requires a single (metaphysically suspect) referent for the DPs in the two expressions. So, the externalist might insist, there need not be some single object that satisfies ‘is bankrupt’ and ‘is a cylinder’, since ‘Hirshhorn-Museum_f’ and ‘Hirshhorn-Museum_b’ denote different objects.

But this reply will not do. The same speaker that endorses (16) and (17) would also endorse the following:

(19) The Hirshhorn-Museum is bankrupt and it is a cylinder.

Whatever the technical details are that govern anaphoric meanings, ‘it’ in (19) must derive its meaning and referent from ‘Hirshhorn-Museum’. Whichever lexical entry the context might supply (‘Hirshhorn-Museum_f’ or ‘Hirshhorn-Museum_b’) the meaning of the anaphoric ‘it’ is exhausted by the referent of whichever lexical item is demanded by the context. Thus, for the EM theorist to explain how competent speakers treat (16)–(19), there must be some single worldly object that is both a cylinder and bankrupt. So even if we grant the EM theorist his homophonous response, the ontological concern remains.

Pietroski (2005) also notes that while a natural language speaker could endorse expressions like (16)–(19), the sentence in (20) is strange in a way (19) is not:

(20)# The Hirshhorn-Museum is a bankrupt cylinder.

The oddity of (20) in conjunction with the acceptability of (19) (and the plethora of sentences like them) is unexplained by a semantics that treats meanings as determining truth-conditions. On any EM theory, the way the world would have to be in order for (19) to be true would also make (20) true: there is some object, the referent of ‘Hirshhorn-Museum’ that is both bankrupt and is a cylinder.²⁸ Insofar as satisfiers of mutually binding predicates of different ontological types are absent from the domain of worldly objects, this speaks against EM theories of natural language meanings that require them.

But, the Realist might bite this bullet, as some semanticists do (see Ludlow, 2003, 2011). They might just stipulate that the domain contains objects that are at once both concrete and abstract. However, this bullet biting is both unmotivated by the externalist argument, and tastes far worse than the Realist might suspect. The externalist *hypothesis* contends that, given the pedestrian objects of the world like chairs and rabbits—objects that we have good *antecedent* reasons to posit—a theory of meaning can be developed given only this domain of pre-theoretically plausible things. The externalist proposal is that natural language meanings can be rendered comprehensible without appeal to the mysterious existence of things like *Sinne*. Thus, the theory garners intuitive support because we are not forced to accept into our domain a vast hierarchy of bizarre objects, like *Sinne*.

In this light, the Realist’s bullet biting is quite strange.²⁹ As a means of avoiding the troubling consequences of viewing meanings as myste-

²⁸ While concatenating predicates does not always yield an expression with the same meaning as conjoining them, this does not seem like such a case. Consider:

- (1) This is a fake diamond.
- (2) This is fake.
- (3) This is a diamond.

While (1) implies (2), it does not entail (3)—in fact (3) must be false if (1) is true. This and other examples are problematic for straightforward applications of concatenating predicates, but the difficulty here is not an ontological one. The sort of objects that satisfy ‘is a diamond’ can also satisfy ‘is fake’—namely physical objects.

²⁹ Famously, Lewis (1986) argues that our best semantics for modal expressions posits a vast plenitude of extant possible concrete particulars that stand in spatio-temporal relations to form possible worlds. The argument on offer contends that these proliferate posits are necessary because the cost of failing to accept these concrete *possibilia* are outweighed by the theoretical benefits they confer. But, this strategy is often misconstrued. These *possibilia* are not objects we merely discover in the process of analyzing meanings. Our use of modal expressions in natural language reveals a commitment on behalf of the natural language speaker to the existence of *possibilia* (concrete or otherwise) only if one *assumes* an externalist semantics. As Ludlow rightly states,

...ontology is tied to the demands of our scientific theory of the semantics of natural language, and not the kinds of entities and objects that members of a

rious *Sinne*, the Realist accepts the existence of hybrid abstract-concreta (concrete-abstracta?), trading the mysterious for the bizarre. The further point is that in accepting the existence of ontologically bizarre hybrid objects, the Realist is *not* merely acknowledging the ontological entailments of natural language usage. Rather, she is making a prediction about the kinds of things we should expect to find in the domain, if the externalist hypothesis is correct. In the absence of any pre-theoretical evidence that there are such things, she bears the burden of providing a good reason for believing that such things exist, beyond the fact that their existence supports her hypothesis.

Notice too, that the nature of these bizarre objects is distinct from those typically associated with Chomskyan critiques of externalism. Unlike ‘flaws’ (Chomsky 1981), or ‘the average man’ (Hornstein 1984) the objects needed to satisfy expressions like (19) are not simply *abstracta* developed for the purposes of scientific theorizing (Ludlow 2011: 135–136) intricately characterized by novel logics (Kennedy & Stanley 2009). Whatever abstract-concreta are, they are far less familiar than mere abstractions, and should be less palatable to the externalist.

But maybe our dislike for the bizarre is unwarranted, based on some vestigial aspect of our human conceptual machinery. Maybe these abstract-concreta exist, happily residing in the domain and validating the Realist’s predictions. Or maybe hybrid objects are more palpable than *Sinne*.³⁰ The Realist’s perverse predictions, however, do not end there. The externalist hypothesis predicts the existence of other hybrids, including hybrid properties, and relations given the lexical flexibility of expressions like ‘bilingual’, ‘lost’, and ‘defeat’.

In a conversation about the ability to acquire a second natural language, one might use the following English expressions:

particular culture might believe in (Ludlow 2011: 142).

To engage in the latter kind of investigation is to do psychology. But, the Realist is investigating the structure of the world, not the structure of our minds. As such, to indicate that the *theory* demands certain metaphysical posits is a burden the theory has to bear, not an analytic consequence of the fact that linguistic expressions have meanings. If (\mathcal{E}) is true, then there are such things as concrete *possibilia*. In this vein, some argue that we need even more things: situations (Barwise & Perry 1983), perspectives (Schein 2002), modes of presentation (Ludlow 1995), and fictional objects (Thomasson 1999). But we only need these things if the best theory of meaning is externalist—a *fortiori* that the theory requires a domain with such things counts against the merits of the theory. Adding to this list of posits the abstract-concreta required to address lexical flexibility, the class of entities required to support the externalist hypothesis looks less and less like the pedestrian objects of everyday experience. Concomitantly, the externalist hypothesis looks less like a plausible theory that makes use of the everyday objects we are familiar with, and more like a theory with implausible commitments.

³⁰ Notice how this vein of reply assumes a false dichotomy, that one either accepts mysterious, reified meanings as things, or accepts *whatever* things an externalist semantics requires. For other alternatives again see Hinzen (2007); Hinzen (2014); Pietroski (2005); Pietroski (2008); Pietroski (2010); Pietroski (forthcoming).

(21) The child is bilingual.

(22) John is bilingual.

Likewise, one could (roughly) conjoin the meanings of these two expressions into a single sentence using either of the following acceptable expressions:

(23) The child is bilingual and so is John.

(24) The child and John are bilingual.

Of course, an externalist semantics can accommodate the meanings of these expressions, whereby their truth-conditions are satisfied just in case there is a (salient) child, a John, and both of them instantiate the same particular property. To put the matter somewhat formally, (23) and (24) are true just in case:

(24') $\iota x. \exists y. \text{CHILD}(x) \ \& \ \text{JOHN}(y) \ \& \ \text{BILINGUAL}(x) \ \& \ \text{BILINGUAL}(y)$

In the expressions (21)–(24), the expression ‘bilingual’ has a single, univocal meaning, as reflected in the single truth-conditional predicate ‘BILINGUAL’. For the externalist this identifies some *single* property, say the property had by all things that acquire/speak/know two languages. As such the sentences in (21)–(24) are well captured by an externalist semantics.

But ‘bilingual’ is lexically flexible. While walking the streets of London, Ontario, Canada, I came across an empty box outside a franchise sandwich shop. Printed on the outside of this box was the expression ‘bilingual napkins’ which presumably identified the box’s contents as napkins on which information is printed in two different languages. Sitting in this franchise with a group of friends, and noticing the features of one such napkin, a competent speaker could well say to their compatriots the following acceptable sentence:

(25) The napkin is bilingual.

Supposing this group also knew our bilingual friend John, a competent speaker could say, and the group would no doubt accept as true, the following acceptable expression:

(26) The napkin is bilingual and so is John.

However, the following truth-conditions, which for the externalist are determined by the meaning of (26) are *not* satisfied in this situation:

(26') $\iota x. \exists y. \text{NAPKIN}(x) \ \& \ \text{JOHN}(y) \ \& \ \text{BILINGUAL}(x) \ \& \ \text{BILINGUAL}(y)$

Since both bilingual predicates in (26') are derived from the single use of ‘bilingual’ in (26), they must have the same truth-conditions. As such, for the externalist they must pick out the very same property. If that property pertains to the acquisition of multiple languages, the napkin clearly fails to instantiate such a property. And if the property pertains to the kinds of orthography printed on a thing’s exterior, John does not count as having such features. For the externalist, (26) turns out to be false in the situation described—the same situation that makes both (22) and (25) true. Given that (26), on an externalist semantics, has the meaning

attained by conjoining (22) and (25) (as indicated by the relationship between (21)–(23)), externalism should predict that competent speakers accept that (26) is true. That it does not is a problem for EM theories.

As before, the externalist could claim that ‘bilingual’ is homophonous, with two lexical entries ‘bilingual_a’ and ‘bilingual_o’ pertaining to the aforementioned acquisitional and orthographic properties (respectively). But just as in the example with ‘the Hirshhorn Museum’, this reply will not do. Whichever lexical entry the context demands, the truth-conditional predicates in (26’) will have the same satisfaction conditions as each other. The ‘bilingual_a’ interpretation of these predicates leaves the first unsatisfied by *the napkin*, while the ‘bilingual_o’ interpretation leaves the second unsatisfied by *John*. Either way, externalism makes the wrong prediction that, relativized to the situation, (26) is (treated as) false (by competent speakers). The externalist’s remaining response is to accept that there is some single bizarre, hybrid (or multifarious) property that admits to having shifting satisfaction conditions within a single context.

As with the case involving the flexibility of ‘the Hirshhorn-Museum’, the Realist could bite this bullet. She can simply accept that the domain contains not only hybrid abstract-concreta, but also hybrid properties such that the very same property can be instantiated in distinct ways by disparate objects within the same context.³¹ But again, as with ‘the Hirshhorn-Museum’, even this (bizarre) concession fails to explain the distribution of competent speaker judgments. Consider

(26) The napkin is bilingual and so is John.

(27)# The napkin and John are bilingual.

The expression in (27) is strange in a way that (26) is not. This strangeness, whatever it amounts to, seems to be a fact about the meanings of the constituent expressions of the sentence, since there is no general prohibition against joint predication (as exemplified in (24)). The syntactic frames of (23) and (24) are repeated in (26) and (27), yet the latter do not bear the same semantic relations to one another as the former, insofar as they are *not* synonymous (as should be clear by the oddity of (27)).

³¹ To be clear, this is not an injunction against multiple realizability. To take the paradigm case, mental properties like *belief* can be realized in Martian brains just as well as human ones. But what is instantiated in these distinct organisms is (say) an entity that plays a particular functional role in the mind of the organism, and indeed the *same* functional role. The worry here is not that the napkin and John embody different ways of instantiating the same property, but that they instantiate *different* properties. Contrast this difference with the manner in which distinct humans are bilingual. John Kerry and Nicolas Sarkozy are both bilingual (let’s suppose), yet the bilingual property is realized in their person in distinct ways. At the very least they differ regarding the languages in which they are fluent: Kerry is fluent in English and French, while Sarkozy is fluent in French and German. The manner in which Kerry and Sarkozy differ is quite clearly not the manner in which John and the napkin differ regarding bilingualism.

The externalist, and the Realist, in order to explain the distribution of competent speaker judgments for sentences containing nouns like ‘the Hirshhorn Museum’ and predicates like ‘bilingual’ are compelled to accept some rather bizarre entities into the domain of worldly things: hybrid objects that exhibit a concrete-abstract duality, and properties that have varying conditions for instantiation across instances within a single context. Neither of these concessions seems pleasant. Worse yet, conceding in these ways still leave unexplained aspects of competent speakers’ judgments, as can be seen by the contrast between (26) and (27).

Finally, consider the following:

(28) Henry lost his key, his lawsuit, and his job.

One can easily imagine a situation in which (28) is deemed both felicitous and true by competent speakers, whereby Henry’s misplacing his car key made him late to the courthouse, which led to his termination. Yet, Henry stands in very different relations to these three objects, all of which are expressed by the single use of ‘lost’ in the sentence. Henry is no longer in possession of his key, while his lawsuit, as a complex activity he participated in, has met with a particular conclusion. And of course, Henry no longer has a job in that a contract he had with some unmentioned individual is no longer binding. On an externalist account the worldly relation that ‘lost’ denotes requires that all three objects, abstract and concrete alike, must (when paired with *Henry*) satisfy that relation, despite these differences. And as with our previous examples, an appeal to ambiguity is not available to the externalist.

Let’s not forget that examples of this sort are numerous:

(29) The chef’s kitchen ran better than an imported car.

(30) Napoleon’s defeat was worse than Kasparov’s.

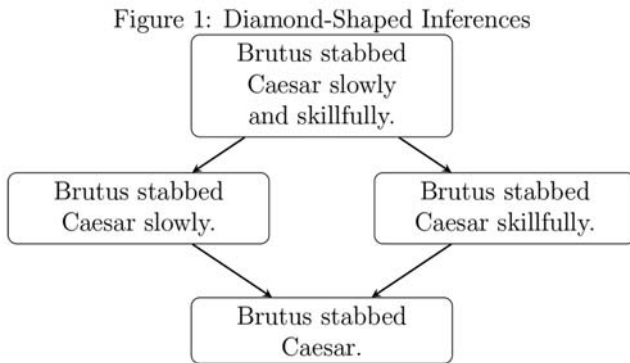
Given the acceptability of sentences like these, expressions like ‘lost’, ‘run’ and ‘defeat’ seem to exhibit lexical flexibility as well. If the Realist is forced to accept the existence of hybrid *relations* as a result of such flexibility, the pedestrian nature of the objects needed to accomplish the externalist’s aims is substantively undermined—especially if this bizarre ontology remains insufficiently explanatory. The point worth underscoring here is that the examples explored in this section are not isolated aberrations in a language that is otherwise well-modeled by externalism. Lexical flexibility is rampant in natural language expressions. To account for the manner in which natural language speakers treat these expressions, the externalist is compelled to accept into her ontology metaphysically bizarre objects, properties, and relations. Rather than viewing these ontological commitments as the price to be paid for an adequate theory of meaning, such requirements might be better seen, or so I contend, as a *reductio* against the externalist hypothesis that requires them.

4.3. An Externalist Reply

The externalist might reply to these worries by leveraging the purported virtues of an externalist semantics. She might contend that even if the flexibility cases rehearsed above require that she bear an ontological burden, such burden bearing is better than the alternative. As a part of an overall externalist theory, the virtues of the theory far outweigh these burdens. One principal virtue that supporters of externalism might trumpet is that understanding meanings in this way uniquely preserves semanticists' main source of linguistic data, the truth-value judgments of competent speakers. To quote a prominent externalist:

In short, intuitions about the truth and falsity of what is said by utterances of sentences have formed the data by which theorists have tested their hypotheses about meaning. There is no other obvious source of native speaker intuitions that are related to meaning. So if we did not have robust intuitions about the truth-conditions of our utterances, it would not be clear how to test such hypotheses; there would be no firm basis on which to construct a theory of meaning. (Stanley 2007: 6)

Consider for example, the landmark insight of Davidson (1967b) in treating the logical form of action sentences as involving quantification over *events*. The sentences in (31) display a particular pattern of inference, as indicated in Figure 1, wherein the arrows represent the direction of inferences that speakers of English are apt to make.



- (31) a. Brutus stabbed Caesar slowly and skillfully.
 b. Brutus stabbed Caesar slowly.
 c. Brutus stabbed Caesar skillfully.
 d. Brutus stabbed Caesar.

Davidson's proposal aims to capture these patterns of inference. Proposing that the logical forms in (31') are indicative of the truth-conditions of the sentences in (31), this approach captures the inferential

judgments of English speakers, by modeling these inferences as logical entailment.

- (31') a. $\exists e[STAB(e, Brutus, Caesar) \ \& \ SLOW(e) \ \& \ SKILL(e)]$
 b. $\exists e[STAB(e, Brutus, Caesar) \ \& \ SLOW(e)]$
 c. $\exists e[STAB(e, Brutus, Caesar) \ \& \ SKILL(e)]$
 d. $\exists e[STAB(e, Brutus, Caesar)]$

Because the proposed logical forms for the expressions in (31) quantify over *events*, the expression in (31'a) entails the other expressions in (31') by way of conjunction reduction. As such, the “diamond-shaped” inference patterns of speakers are captured by an *externalist* theory that takes events as the worldly-satisfiers of expressions.

This reasoning, contends the externalist, only makes sense if the *explanada* of the hypothesis are the inferential judgments of speakers, as judgments about the concomitant truth of collections of sentences. Thus the purported justification for an externalist semantics is that it maintains the theoretical import of speaker judgments. To deny that sentences have externalist meanings is to deny the connection between meaning and truth that renders these judgments worthy of capture. In this manner the externalist might contend, if meanings have nothing to do with truth, then these truth-value judgments are not indicative of expression meanings and of no use for semantic investigation.

However, denying the externalist thesis (\mathcal{E}) does not also require denying that the truth-value judgments of speakers are relevant data for the purposes of semantic theorizing. Externalists are committed to a *particular* relationship between truth and meaning—namely the one codified in (\mathcal{E}). In denying this, a semanticist need not deny that meaning is related to truth. She must simply deny that meanings *determine* truth-conditions. One can hold that natural language speakers can use sentences to make utterances that are true, and still deny the externalist thesis. And this can be done without denying that there is some “systematic” manner in which meaning is related to truth (*pace* Stanley 2007: 8). Such a view merely holds that the systematic manner in which linguistic meaning relates to the external world involves the interaction of multiple non-linguistic cognitive and external systems that connect in complex ways. This complexity might be systematic, but because many of the systems involved are extra-linguistic (and not semantic) the meaning of an expression will not, in the absence of this complex interaction, determine its truth-conditions.

One can, as Stanley does, amalgamate this motley group of disparate non-linguistic systems³² under the term ‘context’. And if by ‘con-

³² The diversity of components that collectively make up the “context” of an utterance, so construed, is important to note. Shared human systems that recognize gaze following, emotional facial gestures, object detection, agency detection, and many others, not to mention the external physical systems that govern “normal” visual and auditory environments, all fall under the “context” that determinately links meanings to truth. But if one wants to know how meanings differentially interact with these various systems in order for a speaker to utter something true,

text’ one includes whatever is needed to derive truth-conditions from meanings, then trivially the meaning of an expression (plus the “context” of its use) determines its truth-conditions. But, the gap between what a sentence means and what a speaker communicates in conveying that meaning *via* a linguistic utterance admits to some marked complexity—as noted in §4.1. Respecting, and not merely masking, the complexity of this relationship between the meaning of a sentence and the truth of an utterance not only preserves the (nuanced) use of truth-value judgments as linguistic data, but it opens up *new* sources of data (Pietroski et al. 2009; Lidz et al. 2011; Vogel et al. 2014). So, far from making the semanticist’s task impossible (or without basis) denying (\mathcal{E}) expands the data-set for the theoretician, while preserving the utility (though augmenting the informativeness) of speaker judgments.

5. *Mental Content*

In the previous two sections we saw arguments that highlight the difficulty in accepting externalist accounts of language (EL views), and externalist theories of meaning (EM views). Arguments against an EL view push an externalist to adopt externalist views about *mental content*. If languages are not external objects, but rather aspects of the human mind/brain, then to rescue the EM view an externalist is committed to the view that the content of mental representations can be characterized externally, as relations between representations (or concepts) and worldly objects. As such, some arguments for internalism address both externalist theories about mental content, and the relationship between mental content and linguistic meaning. In this section I present these concerns.

5.1. *Naturalist Theories of Content*

Both sentences and thoughts seem to be about the world, and thereby exhibit intentionality. The close proximity of these disciplines gives rise to a simple solution to the problem of intentionality for language. Namely, that the problem of intentionality is solved at the level of thought, not language. An enticing view about the relationship between thought and language is that the contents and structure of our thoughts are merely mirrored in language. If the structure of natural language mirrors the structure of thought, wherein an expression in a language is merely a way of making public some particular thought composed of conceptual content, then the intentionality (and meaning) of an expression simply tracks that of the concepts used to compose the expressed thought. On such a view, natural language expressions are merely labels for thoughts, and likewise, words are merely labels

abstracting over these differences by indicating that the context somehow fills this gap is no answer at all. Worse yet, it commits one to a theory of meaning that is thereby incapable of addressing such questions.

for concepts, as a way of making them articulable. Call this the *label theory* of linguistic meaning.³³

So long as the language I speak syntactically composes in a way commensurate with the structure of my thoughts, linguistic meaning would perfectly mirror conceptual content. On this proposal, the meanings of our expressions would hook up with the world *via* conceptual content, so long as conceptual content can be characterized externally. So, to the degree that our concepts align with the “fine structure” of the world, expressions of a natural language will likewise accord with the mind external objects of the world.

But why should one insist that linguistic meaning is mediated by our conceptual system? For one, this answers the problem of intentionality at the level of language. But more importantly for the Realist, the desire here relates to naturalistic explanation. As we saw, a substantive source of contention in thinking about linguistic meaning as externalist relates to the aims of the scientific enterprise of linguistics. Facts about the acquisition and productivity of language in humans deserve explanation, and a theory of meaning ought to add to (or at least make possible) an explanation of these facts. As we saw in §3.1 these considerations strongly suggest that the object of study for a naturalistic investigation of language is in the mind. As such, to the degree one thinks that language and thought are independent, a naturalistic inquiry into these matters will address the way in which these distinct mental faculties interact.

There is no shortage of literature addressing the viability of naturalistic accounts of content.³⁴ Whether or not naturalistic accounts of intentionality are viable is beyond the scope of this work, but the point I want to emphasize here is that the force of the arguments presented so far against externalism compel the externalist to adopt two contentious views: a labeling theory of linguistic meaning, and a naturalistic account of intentionality. I'll not take the time to illustrate the contentiousness of the latter,³⁵ but the former position is worth analyzing, in part because so many philosophers seem to adopt this view without much defense.

³³ Fodor (1975) seems to hold this view. Jackendoff (2002) explicitly adopts this position, though not by this name. Oddly enough he defends an internalist proposal for linguistic meaning on the basis that no naturalistic account of external mental content is plausible.

³⁴ For a good survey see the introduction to Macdonald & Papineau (2006).

³⁵ This has been done by many, and better than I could hope to do here. See Jackendoff (2002); Loewer (1997); Boghossian (1991); Godfrey-Smith (1989); and McGinn (1982).

5.2. Labeling Theory of Meaning

The labeling theory of meaning views the relationship between words and concepts as one of labeling, whereby words are like labels for concepts. My use of ‘cat’ is just a way of tokening the CAT concept in my audience, whose content serves as the meaning of ‘cat’ in any expression that uses it, like ‘The cat is on the mat’. Linguistic meaning on this view is just conceptual content. Such a view (as indicated above) can still be meaningfully externalist, if the contents of our concepts are externalist. This labeling view is pervasive amongst philosophers. In fact, the view is often adopted as obvious, without much need to articulate that indeed adopting the view embodies a collection of commitments about the relationship between the human language faculty and the conceptual system. Burge (1975) is a paradigmatic example. In discussing the expansiveness of his famous ‘arthritis’ case, Burge writes:

On the other hand, the [arthritis] thought experiment does appear to depend on the possibility of someone’s having a propositional attitude despite an incomplete mastery of some *notion* in its content ...Suppose a subject thinks falsely that all swans are white ...that ‘swan’ means ‘white swan’ (Burge 1975: 83) (my emphasis)³⁶

Burge treats the content of a concept, or *notion*, which plays an important role in determining the content of the propositional attitude someone might hold, as no different than the meaning of a natural language word: hence the notion SWAN has the same meaning as ‘swan’. In his book-length critique of Burge’s account of wide-content Segal (2000) commits to this same theory about the relationship of words to concepts.

Zowie and Twin Zowie both say “My engagement ring is studded with diamonds.” Are the concepts expressed by their words “diamond” the same? (Segal, 2000, p. 6)

...

- Let w be the focal word
- Let c be the concept [the subject] expresses by w .

(Segal 2000: 67)

In more contemporary literature, Weber (2005) writes

The meaning of the term “gene” has changed several times in the history of twentieth-century genetics. If we distinguish between a term’s sense and its reference, it is possible that the term’s sense has changed, but not its reference...I have examined both the reference potential and the reference connected with different historical versions of the gene *concept*. (my emphasis) (Weber 2005: 228)

As a final example, consider this passage from Clark & Chalmers (1998), wherein they discuss an opponent’s (possible) response to their thought experiment involving Otto, and his purported belief about the location of a museum:

³⁶ Here Burge uses ‘notion’ as a way of talking about the content of a concept: “Talk of notions is roughly similar to talk of concepts in an informal sense”(Burge 1975: 83).

An opponent might put her foot down and insist that as she uses the term ‘belief’, or perhaps even according to standard usage, Otto simply does not qualify as believing that the museum is on 53rd street. We do not intend to debate standard usage; our broader point is that the *notion* of belief...(my emphasis) (Clark & Chalmers 1998: 14)

Here the slide from the meanings of a linguistic expression ‘belief’ (as evident in its “standard” usage) to the concept (or notion) of belief, is clear, blatant, and offered without explanation or defense.

Linguistic meaning, on this often assumed view, is simply conceptual meaning. But this view has the following consequences: first, the extension of our words must have the same extension as their underlying conceptual meanings, and second, the syntax of natural language must be mirrored in the composition of thought. Neither of these consequences seem well supported by the way natural language speakers treat the meanings of expressions.

The flexibility of natural language expressions speaks against the first consequence.

(19)The Hirshhorn-Museum is bankrupt and it is a cylinder.

If the extension of the concept HIRSHHORN-MUSEUM is to capture the meanings that natural language users apply to the term ‘Hirshhorn-Museum’ then the extension of the concept better include both the concrete building that houses artworks, and the abstract institution that employs hundreds of people. As we’ve seen, many natural language expressions bear meanings that do not track the domain of objects in this way. Thus whatever thought corresponds to (19), and thereby stands as the meaning of (19), it must either treat ‘Hirshhorn-Museum’ as labeling two distinct concepts, or have a content such that some (abstract) object (or some building) is both cylindrical and bankrupt. The former avenue belies the manifest relationship between the uses of the English term, and fails to account for the felt relatedness of these uses. The latter option has much more bizarre metaphysical commitments, since we have no other reason (other than a commitment to particular views about semantics) to postulate such an entity.

Turning to the second consequence, if language mirrored the structure of thought, then thoughts should compose much the way expressions do. More strictly, the meanings of linguistic expressions and their underlying logical forms should mirror the structure of the concepts those forms express. The deep structural syntactic frames that make up interpretable expressions in a natural language must mirror the structure of their underlying concepts. In this vein, consider the following sentence:

(32)Wilbur kicked Fred.

Paying attention to the syntax of this construction, and adopting the labeling theory commitment, we ought to conclude that the KICK concept is dyadic. The word ‘kicked’ in the complete expression in (32) takes a subject and an object, and likewise we would expect the related concept

to take two elements to form a complete thought. Thus the thought expressed must make use of a concept like:

(33) KICK(*_s_ _o*)

which when saturated with two elements, makes the complete thought

(34) KICK(*Wilbur, Fred*)

However, if this dyadic notion of KICK is supposed to underlie all meaningful uses of 'kick', as implied by the labeling theory, the following expression is an apparent counterexample:

(35) Wilbur kicked Fred with his foot.

Given the syntactic structure of the expression in (35), the underlying conceptual meaning must have a *triadic* structure, to make room for the instrument used in the kicking:

(36) KICK(*_s_ _o _i*)

For any way of differentiating concepts, surely addicity falls under the identity condition for a given concept. That is, concepts with different addicities must be *different* concepts. So, (33) and (36) cannot be the same concept. As such, the meaning of 'kicked' in (32) and (35) is different on the labeling view, insofar as 'kicked' labels concepts that are (of) different (addicities). This entailment leaves unexplained why a competent speaker would find that both (32) and (35) are felicitous descriptions of the same kicking.

A defender of the labeling view might hold that really we have only one KICK concept, with sufficient addicity to accommodate all uses of 'kick', and thereby holding their meaning constant across various uses. As such, the concept in (36) is the only KICK concept, made use of in expressions where the instrument of the kicking is unmentioned.

There are three problems with that response: first, this requires that many expressions that make use of the transitive 'kick' have implicit content of an unspoken instrument. And there seems to be no syntactic evidence that such expressions have any such implicit content. Second, such a triadic concept will not capture the meanings of expressions like

(37) Wilbur kicked Fred the ball.

(38) Wilbur kicked Fred the ball with his toe.

The underlying conceptual meaning for (38) must have a *tetradic* addicity, to make room for the indirect and direct objects in the syntactic structure of the expression. Insisting here that the single conceptual meaning for 'kick' is a *tetradic* concept is implausible. While one might entertain the plausibility that transitive uses of 'kick' leave some unspoken implicit content about what instrument was used in a given kicking, surely such uses do not leave the existence of (nonexistent??) indirect objects implicit, as would be required if (39) was the underlying conceptual meaning of (32').

(39) KICK(*WILBURs, FREDdo, NOTHINGo, FOOTi*)

(32') Wilbur kicked [nothing to] Fred [with his foot].

And plainly, (32) does not mean what (32') means.

The third problem for the super-addicity move pushes in the opposite direction. Sentences like (40) seem to require conceptual meanings that are monadic.

(40) Wilbur kicked.

Insisting here that the expression in (40) really contains implicit content that reflects the underlying tetradic conceptual structure in (39) strains good explanation.

Of course, the label theorist could respond to this data by treating all this as evidence that there are really multiple words 'kick' with multiple KICK concepts as their meanings. There are, on this reply, multiple homophones 'kick' each paired with a different concept depending on whether they have direct objects, instruments, and/or indirect objects. However, such a response treats the difference between these uses of 'kick' like the difference between 'kick' and 'punch'—they are different words, with different conceptual meanings. This of course leaves unexplained what is obvious, that the many uses of 'kick' describe quite similar actions, and are conceptually related. The events these various uses of 'kick' describe bear striking features in common—those features that make them plausible kickings in the first place. A theory about the relationship between the meaning of words and the content of concepts that leaves such basic facts unexplained is troubled.

The point then is this: the relationship between linguistic meaning and conceptual content is not nearly as simple as the labeling theory would have it, as the case of (the addicity of) 'kick' and KICK shows. And addicity is but one feature of the relationship between lexical items and concepts that admits to some *prima facie* complexity.³⁷ For the externalist, this should be troubling news, since this means that the path from the meaning of a term, through the content of the associated concept, to its worldly extension is rather complex. In short, the IL-EM

³⁷ Other aspects of the content of our concepts, apart from their structure, highlight the complex connection between words and the concepts that underwrite them. Consider the different ways in which the predicate 'is blue' applies to objects, and what this says about the complex application of the blue concept.

- (1) The house is blue.
- (2) The marker is blue.
- (3) The iris is blue.
- (4) The sky is blue.

The truth-conditions for the color predicate in (1) that would make it true of some house, would not, when applied to some marker, make (2) true, despite the fact that they appear in the same syntactic frame. So if the meaning of expressions are a result of the satisfaction conditions of their underlying concepts, the satisfaction conditions for the thoughts BLUE(HOUSE) and BLUE(MARKER) are not merely going to be attributable to the differences in the extension of HOUSE and MARKER—*mutatis mutandis* for (3). And while many uses of (4) are considered true by competent speakers, what is far from clear is which object is picked out such that it satisfies 'is blue' in any of the ways just mentioned here.

theorist is committed to two views about language and content that are contentious, one of which (given our discussion here) seems implausible. Not only are such theorists saddled with giving a sufficiently plausible *naturalistic* account of mental content, they are also saddled with the troubles articulated here for the label theory of linguistic meaning.

6. *Natural Language and Ontology*

The arguments thus far have been multifaceted, but direct. What they have shown is that externalist proposals about the nature of languages as objects, and about linguistic meaning—in particular (\mathcal{E})—are difficult theses to defend. In this closing section I'll indicate how this difficulty undermines the Realist's metaphysical methodology. Primarily, if (\mathcal{E}) is dubious, then appealing to the truth-conditions of natural language expressions, or the satisfaction conditions of their purported conceptual meanings, as justification for metaphysical conclusions is without foundation.³⁸

The problems detailed thus far for (\mathcal{E}) undermine the fruitfulness of the Realist's default metaphysical methodology.³⁹ Ontological investigation proceeds by analyzing natural language usage. Roughly, the Realist makes use of her competence with a given natural language, since such competence ensures that she understands the meanings of natural language expressions. Under the guise that such meanings are externalist, she derives from them the ontology one is committed to in accepting the truth of a given expression. Taken together, the ontological commitments of all the true sentences determine what there is. In the opening sections of this paper I sketched the Realist's methodology.

³⁸ This is especially true if the human ability to construct complex thoughts from different conceptual domains depends on the human language faculty (Carruthers 2002; Spelke 2003; Jackendoff 1990, 1996; Bloom 2000). Even if we grant that our concepts have satisfaction conditions that accord with the structure of reality, once those concepts are put to work by the language faculty in building meanings to sentences, there's no assurance that the content of the resulting construction will retain such a tight connection to the world (see Glanzberg 2011). And of course there is no guarantee that our concepts accord with reality's structure. The work of Michotte (1946 [1963]) illustrates the difficulty of such certainty with regard to our judgments of causation, where clearly non-causal scenes are judged by subjects as exhibiting causation. *A fortiori* these judgments persist even when objects interact in ways nearly identical to clearly non-causal events (Scholl & Nakayama 2002).

³⁹ Such a position is Realist because it holds that there is an objective structure to the world. However, the view is also importantly Quinean, in the sense that it adopts a methodology of looking to language, and to quantificational structure in particular, to settle ontological disputes. While these two positions seem interrelated, there is reason to think neither entails the other (Hirsch 2002). The arguments outlined here do not undermine Realism *tout court*, even if they directly undermine the Quinean Realist position adopted by many philosophers and metaphysicians. For these reasons I've highlighted the distinction here, though throughout the paper, I use 'Realist' and its neighboring forms as short hand for the more cumbersome and possibly misleading 'Quinean Realist'

I'll close here with a restatement of that widely-adopted strategy and articulate the reasons to reconsider its merits.

In §1 we saw that the Realist holds that there is a unique language (an interpretation of the existential quantifier) whose quantificational structure mirrors the structure of reality.⁴⁰ Put more explicitly, Realist

...inquiry will be guided by ...[an] assumption [that] modern logic's quantificational apparatus mirrors the structure of reality: I assume an ontology of *things*. Moreover, I assume that there is a single, objective, correct account of what things there are. (Sider 2002: xvi)

This account is captured by the meanings of expressions in (what I'll call) the Language of Ontology, or \mathcal{L}_O . With this privileged language in hand metaphysicians can proceed to answer ontological questions by investigating the meanings of expressions in that privileged language, which can be given by way of Tarskian satisfaction by sequences of domain objects. Thus the Realist holds, the objects required to account for the meanings of the true expressions in \mathcal{L}_O are the objects of reality, since this privileged language mirrors reality's (object-based) structure.

To highlight an example of this strategy in action, consider the following points made by Sider (2002) in arguing against certain conceptions of time:

The status of tense is a second issue in the philosophy of time. Tensed sentences are those which presuppose a certain position or vantage point within the whole of time, for example:

It is *now* raining.

It *was* the case that there existed dinosaurs.

I *will* one day visit Utah. (Sider 2002: 12)

In arguing against a presentist theory of time, Sider contends that the presentist cannot clearly account for the truth of sentences that (seemingly) refer to the non-present. Insofar as the presentist denies that there are any ontologically real past or future times, *any* sentence that requires the existence of past/future times must thereby be either meaningless, or simply false. Such sentences have no truth-makers given the presentist's ontology, and thus the presentist cannot account for the truth of tensed sentences.

The success of this argument clearly presumes that the meanings of these natural language expressions determine their truth-conditions, insofar as their meaningfulness depends on the existence of past/future times. The presentist denies that there are past/future times. If the sentences Sider presents are meaningful—which they surely are, given that competent speakers of English have no problem understanding

⁴⁰ Sider expresses the same Realist commitment with different, less perspicuous language elsewhere:

Clearly there are multiple (inferentially and materially adequate) interpretations of quantifiers. As I see it, the real issue is whether any of these interpretations is metaphysically distinguished, whether any of them uniquely matches the structure of the world, whether any carves nature at the joints better than the others. (Sider 2009: 392)

them—then *ex(ternalist) hypothesi* they have truth-conditions. Those conditions are only satisfied if there is some future time where Sider is in Utah, and some past time where dinosaurs are alive and well. The presentist, contends Sider, must admit then that all tensed sentences are false or meaningless, since they have no temporal satisfiers. This consequence thereby seems bad for the presentist.

Of course, a key step in this line of argumentation asserts the truth of (E). As we have seen, this externalist hypothesis is troubled. If the meaning of tensed terms do not determine whether or not they refer to *times* (*pace* externalism), then the move from linguistic meaning to ontological commitment is without warrant. And the supposition, if not false, is (at least) difficult to defend in light of the flexibility of natural language.

However, the Realist has a ready (and plausible) response to this objection. After all, natural languages like English are awash with vexing semantic properties like vagueness, ambiguity, and (apparently) lexical flexibility. As such, there is little surprise that they are ill-suited for the purposes of ontological investigation. The language the Realist needs is one that conforms to the features of classical logic, and none of these semantic properties are tolerated by such logics. But some languages are not deficient in these ways—namely the languages invented in the process of scientific inquiry. On this reply, the privileged language \mathcal{L}_O needed for ontological investigation is the one preferred by our best sciences. After all, scientific inquiry is guided by the expressed purpose of perspicuously describing the world. This process involves making decisions about what terms to use. The results of such inquiry are languages that embody the kind of precision that natural languages like English lack. This embodiment makes these scientific languages better suited for ontological investigation, and thereby better candidates for \mathcal{L}_O .

Sider himself indicates as much:

I hold that the fundamental is determinate...First, no special-purpose vocabulary that is distinctive of indeterminacy ...carves [nature] at the joints. Second, fundamental languages obey classical logic. (Sider, 2011, p. 137)

The fundamental structure of reality does not admit to the fuzziness typical of natural language meanings. Likewise, fundamental languages, those that cleave to the structure of the world, are free of such properties. Their constituents have determinate meanings/referents. The privileged \mathcal{L}_O , the language that mirrors the fundamental structure of reality, should not admit to the kind of indeterminacy we find with natural languages.⁴¹ Instead, this privileged language should have the

⁴¹ Technically, \mathcal{L}_O must typify a class of languages, not a single language. Suppose \mathcal{L}_O contains the terms of imperial measurement. Such a language can be translated into one that uses a metric system. Neither language would cleave to realities structure any better than the other. *Mutadis mutundis* for the many logical operators. Sider's focus is on the interpretation of the quantifiers, and in particular, the existential quantifier. On this point, all the various languages in the \mathcal{L}_O class would agree, for the Realist.

semantics of a Tarskian logic, whereby domain objects constitute the meanings of its expressions.

According to Sider, the process of scientific inquiry yields languages of this ontologically determining sort:

We should believe generally what good theories say; so if a good theory makes an ontological claim, we should believe it. The ontological claim took part in a theoretical success, and therefore inherits a borrowed luster... [But] the conceptual decisions ...also took part in a theoretical success, and also inherit a borrowed luster. (Sider 2011: 12)

Sider suggests that the languages used to express our best scientific theories are better suited for questions of ontology, and that certain sciences do this better than others (Sider 2011: 6). The substantive assumption endorsed by this Realist is that scientific methodology is of sufficient epistemic heft that the languages our best sciences construct are those that match the quantificational structure of the world. And because the languages used to state our best scientific theories are designed to perspicaciously describe the world, the epistemic credentials of naturalistic inquiry assures us that such languages make use of externalist meanings. The value of this move to such languages is measured by the degree to which scientific methods deliver languages that mirror the structure of reality. I've argued elsewhere that this move might be suspect, as terms in core sciences seem to exhibit lexical flexibility in much the way natural languages do (Vogel, under review). Nonetheless, the Realist that embraces this reply must thereby abandon the use of natural languages for the purpose of ontological investigation. The very insistence that there is a distinction between natural and scientific languages (in terms of their ontological credentials) calls into question the use of natural language speaker judgments to adjudicate ontological matters.

It's worth noting how impactful this Realist retreat to scientific languages is. Much of philosophical discourse in both metaphysics proper, and in other philosophical domains where ontological questions seem to matter, relies on the use of natural language expressions, and the purported ontological commitments of speakers that endorse those expressions as true. The example above with tensed sentences is commonplace. Stotz et al. (2004) summarizes the sort of strategy philosophers deploy in pursuing their metaphysical projects:

[In analyzing concepts] [t]raditionally, philosophers have relied on their individual linguistic competence with the corresponding words. When analyzing a concept, the philosopher treats him or herself as a sociolinguistic 'sample of one' Stotz et al. (2004).

This kind of *Conceptual Analysis* tends to proceed by offering up a short *natural language* description of a situation, and then probes whether or not the situation that meets the truth-conditions of the description also serves as a satisfier of the term under dispute. To take an example of this method, consider this case from Lewis (2000) in his discussion about the metaphysics of causation:

[ROCKS]

Billy and Suzy throw rocks at bottles. Suzy throws first, or maybe throws harder. Her rock arrives first. The bottle shatters. When Billy's rock gets to where the bottle used to be, there is nothing but flying shards of glass. [...]
So Suzy's throw causes the shattering. Billy's doesn't. (Lewis 2000: 184)

The familiar method used here is to present a case which makes use of the metaphysical notion in question, and in light of the readers' comprehension of the passage, leverage their intuitive judgments about the described case with regard to that notion. Here, the case is presented to show a flaw in a simple counterfactual notion of causation, and lend support to an ancestral-counterfactual account. The judgments of natural language speakers, namely that the expression 'Billy caused the bottle to break' is false, plays an evidential role in Lewis' argument. This judgment is offered as evidence that the counterfactual conception of *causation*⁴² is troubled. The supposition is that the same situation which makes the claims in ROCKS true, makes the causal claim about Billy and his rock false. As such, the *causal structure* of the situation is not captured by a theory that belies this speaker judgment—a judgment that can only serve as evidence about *causation* if words like 'cause' refer to elements of causal structure. To assume that words so refer is just to accept the troubled externalist hypothesis (E).

Importantly Conceptual Analysis is not of a kind with a naturalistic approach to language. As Chomsky (1965) describes the related methodology for linguistic inquiry, subjects' judgments of the acceptability of a sentence need to be captured by a theory of grammar, with the hypothesis being that the acceptability of those judgments is *explained* by violations of grammatical rules. This makes sense if the *explanandum* is an aspect of the human mind/brain, since the judgments of natural language speakers are (hypothesized by the Chomskyan to be) the product of a mental faculty which includes algorithmic rules for constructing sentence—a grammar. Analogous inquiry in semantics tends to pertain to judgments about entailment, the felicity of a series of descriptions to a scene, or judgments about whether expression pairs have similar meanings. What linguists do not do is introspect into the meaning of words in a language, and certainly not with an eye toward answering ontological questions. Such a methodology plainly makes little sense if one's goal is to describe anything other than the way a particular subject views the world. Only with the added externalist

⁴² Importantly Lewis takes cases like this to illuminate the nature of causation, *not* the semantics for the English expression 'cause', as exhibited by the judgments of competent English-speakers (Lewis 1973). He addresses this question in a footnote, indicating that his proposal regards causal *facts*, *not* linguistic objects. Further, Collins et. al. echo this goal in their introductory contribution to a prominent volume on the metaphysics of causation. In fact, they indicate that the central misstep of a competing analysis defended by Davidson (1967a) is the focus on sentences instead of propositions (Collins et al. 2004: 17). They insist that the evidence brought to bear by dissecting cases, in the manner above, informs us about *propositions* and causal *facts*, not merely linguistic expressions that invoke the term 'cause'.

assumption that the meanings of expressions have real-world denotational meanings could such a methodology sensibly be applied for doing metaphysics. But it is precisely that externalist assumption that I've argued is problematic.

Nonetheless, debates about mental content (Burge 1975, 1979), the ontology of minds (Clark & Chalmers 1998), persons (Parfit 1984), causation (Lewis 1973; Collins et al. 2004), identity (Black 1952), modality (Plantinga 2003), rationality (Williams 1979), moral theory (Foot 1967), and many others require that natural language intuitions play a profound evidential role in settling ontological questions. But if natural language meanings fail to determine the truth-conditions for the expressions they serve, then the ontological commitments of speakers that endorse those expressions as true are (at best) indeterminate. Conceptual Analysis is grounded on the presumption that the truth-value judgments of competent speakers of a natural language in which the proposed thought experiment is written have ontological commitments. The Realist retreat to scientific languages acknowledges the deficiencies of natural languages for this purpose. Such a Realist, in the absence of an adequate response to the worries presented here regarding (\mathcal{E}), should abandon Conceptual Analysis as a method of ontological investigation.

The arguments presented do not rule out the Realist's methodology full stop. A Realist might well respond to these worries by abandoning their use of natural languages for languages more amenable to externalist treatment. In particular, they might adopt the suggestion made by Quine (1960: 221) and defended by Sider (2011), and look to the invented languages used to express scientific theories—languages constructed to avoid the pitfalls of natural language.⁴³ That is, the Realist might contend that while the meanings of natural language expressions do not determine their referents, terms in a scientific language do, insofar as such languages do not suffer from the vagueness and flexibility of natural language expressions.⁴⁴ For this Realist, the privileged language of ontology (\mathcal{L}_O) is a formal language, developed to express our best scientific theories. However, given both the naturalistic commitments of the Realist, and the flexibility natural languages exhibit, the arguments presented here suggest that \mathcal{L}_O cannot be a natural language, or some regimented variant of one. Consequently, the metaphysical methodologies that assume the externalist hypothesis for natural languages, like Conceptual Analysis, are without foundation, and should be abandoned as means of settling ontological disputes.

⁴³ Hence the meaning at the heart of Quine's proclamation "Language is conceived in sin and science is its redemption" (Quine 1973: 68)

⁴⁴ Though this contention too might seem dubious when one considers the various uses of biological terms like 'gene' (Weber 2005; Beurton et al. 2000; Stotz & Griffiths 2004; Stotz et al. 2004; Wilson et al. 2007).

References

- Austin, J. L. 1962. *How to Do Things with Words*. Oxford: Oxford University Press.
- Bach, K. 1999. "The semantic-pragmatic distinction: What it is and why it matters." In K. Turner (Ed.) *The Semantic-Pragmatics Interface from Different Points of View*. Oxford: Elsevier: 65–84.
- Bach, K. 2004. "Minding the gap." In *The Semantics/Pragmatics Distinction*. Stanford: CSLI Publications: 27–43.
- Baker, M. 2002. *The Atoms of Language*. New York: Basic Books.
- Barwise, J., & Perry, J. 1983. *Situations and Attitudes*. Cambridge, Mass.: Bradford Books, MIT Press.
- Beurton, P. J., Falk, R., & Rheinberger, H.J. 2000. *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives*. Cambridge: Cambridge University Press.
- Black, M. 1952. "The identity of indiscernibles." *Mind* 61 (242): 153–164.
- Bloom, P. 2000. *How Children Learn the Meaning of Words*. Cambridge, Mass.: MIT Press.
- Boeckx, C. 2006. *Linguistic Minimalism : Origins, Concepts, Methods, and Aims*. Oxford: Oxford University Press.
- Boghossian, P. A. 1991. "Naturalizing content." In B. M. Loewer, & G. Rey (Eds.) *Meaning in Mind: Fodor and His Critics*. Oxford: Blackwell.
- Bromberger, S., & Halle, M. 1995. "The ontology of phonology." *Studies in Logic and the Foundations of Mathematics* 134: 725–743.
- Burge, T. 1975. "On knowledge and convention." *Philosophical Review* 84 (2): 249–255.
- Burge, T. 1979. "Individualism and the mental." *Midwest Studies in Philosophy* 4 (1): 73–122.
- Cappelen, H., & Lepore, E. 2005. *Insensitive Semantics: A Defence of Semantic Minimalism and Speech Act Pluralism*. Oxford: Blackwell Publishing.
- Carey, S. 2009. *The Origin of Concepts*. Oxford: Oxford University Press.
- Carruthers, P. 2002. "The cognitive functions of language." *Behavior and Brain Sciences* 25: 261–316.
- Chomsky, N. 1965. *Aspects of a Theory of Syntax*. Cambridge, Mass.: MIT Press.
- Chomsky, N. 1975. *Reflections on Language*, chap. The Object of Inquiry. Pantheon Press.
- Chomsky, N. 1977. *Essays on Form and Interpretation*. Oxford: Elsevier.
- Chomsky, N. 1981. *Lectures on Government and Binding* Dordrecht: Foris.
- Chomsky, N. 1986. *Knowledge of Language*. New York: Praeger Publishing.
- Chomsky, N. 2000. *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.
- Clark, A., & Chalmers, D. 1998. "The extended mind." *Analysis* 58 (1): 7–19.
- Collins, J., Hall, N., & Paul, L. A. 2004. "Counterfactuals and causation: History, problems, and prospects." In J. Collins, N. Hall, & L. Paul (Eds.) *Causation and Counterfactuals*. Cambridge, Mass.: MIT Press: 1–57
- Davidson, D. 1967a. "Causal relations." *Journal of Philosophy* 64 (21): 691–703.

- Davidson, D. 1967b. "The logical form of action sentences." In N. Rescher (Ed.) *The Logic of Decision and Action*. Pittsburgh: University of Pittsburgh Press.
- Dummett, M. 1978. "What do I know when I know a language." In Dummett (1993). Oxford: Oxford University Press.
- Dummett, M. 1993. *The Seas of Language*. Oxford: Oxford University Press.
- Fodor, J. A. 1975. *The Language of Thought*. Cambridge, Mass.: Harvard University Press.
- Foot, P. 1967. "The problem of abortion and the doctrine of double effect." *Oxford Review* 5: 5–15.
- girontravel.se 2013. The town of Kiruna. <http://girontravel.se/en/fun-facts/the-town-of-kiruna>
- Glanzberg, M. 2011. "Meaning, concepts, and the lexicon." *Croatian Journal of Philosophy* 11 (1): 1–29.
- Godfrey-Smith, P. 1989. "Misinformation." *Canadian Journal of Philosophy* 19 (4): 533–50.
- Grice, H. P. 1957 [1989]. "Meaning." In H. P. Grice (Ed.) *Studies in the Way of Words*. Cambridge, Mass.: Harvard University Press: 213–223.
- Hinzen, W. 2006. "Internalism and truth." *Mind & Society* 5: 139–166.
- Hinzen, W. 2007. *An Essay on Names and Truth*. Oxford: Oxford University Press.
- Hinzen, W. 2014. "Recursion and truth." In T. Roeper, & M. Speas (Eds.) *Recursion: Complexity in Cognition*, vol. 43 of *Studies in Theoretical Psycholinguistics*. Springer International Publishing: 113–137.
- Hirsch, E. 2002. "Quantifier variance and realism." *Philosophical Issues* 12 (1): 51–73.
- Hornstein, N. 1984. *Logic as Grammar*. Cambridge, Mass.: MIT Press.
- Jackendoff, R. 1990. *Semantic Structures*. Cambridge, Mass.: MIT Press.
- Jackendoff, R. 1996. "Semantics and cognition." In S. Lappin (Ed.) *The Handbook of Contemporary Semantic Theory*. Oxford: Blackwell Publishing: 539–559.
- Jackendoff, R. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
- Kaplan, D. 1977 [1989]. "Demonstratives". In J. Almog, J. Perry, & H. Wettstein (Eds.) *Themes from Kaplan*. Oxford: Oxford University Press: 481–563
- Kennedy, C., & Stanley, J. 2009. "On 'average'." *Mind* 118 (471): 583–646.
- Lewis, D. 1970. "General semantics." *Synthese* 22: 18–67.
- Lewis, D. 1973. "Causation". *Journal of Philosophy* 70 (17): 556–567.
- Lewis, D. 1975. "Languages and language." In K. Gunderson (Ed.) *Language, Mind, and Knowledge*. University of Minnesota Press.
- Lewis, D. 1986. *On the Plurality of Worlds*. Oxford: Blackwell Publishing.
- Lewis, D. 2000. "Causation as influence." *Journal of Philosophy* 97 (4): 182–197.
- Lidz, J., Pietroski, P., Hunter, T., & Halberda, J. 2011. "Interface transparency and the psychosemantics of most." *Natural Language Semantics* 19 (3): 227–256.
- Loewer, B. M. 1997. "A guide to naturalizing semantics." In C. Wright, & B. Hale (Eds.) *A Companion to the Philosophy of Language*. Oxford: Blackwell: 108–126.

- Ludlow, P. 1995. "Logical form and the hidden-indexical theory: A reply to Schiffer." *The Journal of Philosophy* 92 (2): 102–107.
- Ludlow, P. 2003. "Referential semantics for i-languages?" In L. M. Antony, & N. Hornstein (Eds.) *Chomsky and His Critics*. Wiley Online Library: 140–161.
- Ludlow, P. 2011. *The Philosophy of Generative Linguistics*. Oxford: Oxford University Press.
- Macdonald, G., & Papineau, D. (Eds.) 2006. *Teleosemantics*. Oxford: Oxford University Press.
- McGinn, C. 1982. "The structure of content." In A. Woodfield (Ed.) *Thought and Object*. Oxford: Oxford University Press.
- Michotte, A. 1946 [1963]. *The Perception of Causality*. New York: Basic Books.
- Parfit, D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Piattelli-Palmarini, M., Uriagereka, J., & Salaburu, P. 2009. *Of Minds and Language: A Dialogue with Noam Chomsky in the Basque Country*. Oxford: Oxford University Press.
- Pietroski, P. M. 2005. "Meaning before truth." In G. Preyer, & G. Peter (Eds.) *Contextualism in Philosophy*. Oxford: Oxford University Press: 255–302.
- Pietroski, P. M. 2008. "Minimalist meaning, internalist interpretation." *Biolinguistics* 4: 317–341.
- Pietroski, P. M. 2010. "Concepts, meaning and truth." *Mind & Language* 24 (3): 247–278.
- Pietroski, P. M. forthcoming. *Conjoining Meanings: Semantics Without Truth-Values*. Oxford: Oxford University Press.
- Pietroski, P. M., Lidz, J., Hunter, T., & Halberda, J. 2009. "The meaning of 'most': Semantics, numerosity and psychology." *Mind & Language* 24 (5): 554–585.
- Plantinga, A. 2003. *Essays in the Metaphysics of Modality*. Oxford: Oxford University Press.
- Quine, W. 1960. *Word and Object*. Cambridge, Mass.: MIT Press.
- Quine, W. 1973. *The Roots of Reference*. La Salle: Open Court.
- Rolander, N. 2013. "Swedish city of Kiruna plans massive relocation." *Wall Street Journal*. <http://on.wsj.com/1w3GxiL>
- Rothschild, D., & Segal, G. 2009. "Indexical predicates." *Mind & Language* 24 (4): 467–493.
- Russell, B. 1957. "Mr. Strawson on referring." *Mind* 66 (263): 385–389.
- Schein, B. 2002. "Events and the semantic content of thematic relations." In G. Preyer, & G. Peter (Eds.) *Logical Form and Language*. Oxford: Oxford University Press: 263–344.
- Scholl, B. J., & Nakayama, K. 2002. "Causal capture: Contextual effects on the perception of collision events." *Psychological Science* 13 (6): 493–498.
- Segal, G. 2000. *A Slim Book on Narrow Content*. Cambridge, Mass.: Bradford Books, MIT Press.
- Sider, T. 2002. *Four Dimensionalism: An Ontology of Persistence and Time*. Oxford: Clarendon Press.
- Sider, T. 2009. "Ontological realism." In D. Chalmers, D. Manley, & R. Wasserman (Eds.) *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press.

- Sider, T. 2011. *Writing the Book of the World*. Oxford: Oxford University Press.
- Spelke, E. S. 2003. "What makes us smart? Core knowledge and natural language." In *Language in Mind: Advances in the Study of Language and Thought*. Cambridge, Mass.: MIT Press: 277–311.
- Stainton, R. 2006. "Meaning and reference: Chomskian themes." In E. LePore, & B. Smith (Eds.) *The Oxford Handbook of Philosophy of Language*. Oxford: Oxford University Press: 913–940.
- Stanley, J. 2000. "Context and logical form." *Linguistics and Philosophy* 23 (4): 391–434.
- Stanley, J. 2007. *Language in Context: Selected Essays*. Oxford: Oxford University Press.
- Stotz, K., & Griffiths, P. 2004. "Genes: Philosophical analyses put to the test." *History and Philosophy of the Life Sciences* 26 (1): 5.
- Stotz, K., Griffiths, P. E., & Knight, R. 2004. "How biologists conceptualize genes: An empirical study." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 35 (4): 647–673.
- Strawson, P. F. 1950. "On referring." *Mind* 59 (235): 320–344.
- Thomasson, A. L. 1999. *Fiction and Metaphysics*. Cambridge: Cambridge University Press.
- Vogel, C. A. under review. "Realism and lexical flexibility".
- Vogel, C. A., Wellwood, A., Dudley, R., & Ritchie, J. B. 2014. "Talking about causing events." *Baltic International Yearbook of Cognition, Logic and Communication* 9.
- Weber, M. 2005. *Philosophy of Experimental Biology*. Cambridge: Cambridge University Press.
- Williams, B. 1979. "Internal and external reasons." In R. Harrison (Ed.) *Rational Action*. Cambridge: Cambridge University Press: 101–113.
- Wilson, R. A., Barker, M. J., & Brigandt, I. 2007. "When traditional essentialism fails: Biological natural kinds." *Philosophical Topics* 35 (1): 189.
- Wittgenstein, L. 1953. *Philosophical Investigations*. MacMillan.
- Wittgenstein, L. 1972. *On Certainty*. Harper Perennial.

Socio-historical Causal Descriptivism. A Hybrid and Alternative Theory of Names

CHEN BO

Department of Philosophy, Peking University, PR China

This paper argues for a hybrid and alternative theory of names—Socio-historical Causal Descriptivism, which consists of six claims: (1) the referring relation between a name and an object originates from a generalized “initial baptism” of that object. (2) The causal chain of the name N firstly and mainly transmits informative descriptions of N’s bearer. (3) The meaning of N consists of an open-ended collection of informative descriptions of N’s bearer acknowledged by a linguistic community. (4) With respect to practical needs of agents there is a weighted order in the collection of descriptions of N’s bearer. (5) The meaning or even partial meaning of N, together with the background of a discourse, the network of knowledge, speaker’s intention, etc., determines the referent of N. (6) All names have their own referents, including physical individuals, and parasitic, fictional, or intensional objects; there are few names absolutely without reference.

Keywords: Name, description, descriptivism, referentialism, socio-historical causal descriptivism.

In order to answer the questions “how does language work?” and “where does linguistic meaning come from?”, I argued for Social Constructivism of Language and Meaning (SCLM for short) in another paper. SCLM consists of six theses: (1) The primary function of language is communication rather than representation, so language is essentially a social phenomenon. (2) Linguistic meaning originates from the causal interaction of humans with the world, and from the social interaction of people with people. (3) Linguistic meaning consists in the correlation of language to the world established by collective intentions of a language community. (4) Linguistic meaning is based on the conventions set up by a language community in their long process of communication. (5) Semantic knowledge is empirical and encyclopedic knowledge

condensed and distilled, and the uses of language accepted by a linguistic community. (6) Language and meaning change rapidly or slowly as the communicative practice of a linguistic community does. The crucial point of SCLM is to focus on the triadic relation among language, humans (a linguistic community) and the world, rather than the dyadic relation between language and the world (cf. Chen Bo 2015: 87).

In this paper, by an appeal to SCLM, I will review the “war” between descriptivism and referentialism in contemporary philosophy of language (cf. Lowe 2007: 27), and argue for a hybrid but still alternative theory of names—I call it “Socio-historical Causal Descriptivism” (SHCD for short). SHCD aims to answer the question of how people use names, especially proper names, to refer to their referents in natural language, and it contains other six claims: (1) The referring relation between a name and an object originates from a generalized “initial baptism” of the object. (2) The causal chain of name *N* transmits informative descriptions of *N*’s bearer. (3) The meaning¹ of *N* is an open-ended collection of informative descriptions of *N*’s bearer acknowledged by a linguistic community. (4) With respect to the practical needs of agents there is a weighted order in the collection of descriptions of *N*’s bearer. (5) The meaning or even partial meaning of *N*, together with the background of a discourse, the network of knowledge, speaker’s intention, etc., determines the referent of *N*. (6) All names have their own referents, including physical individuals, and parasitic, fictional, or intentional objects; there are few names absolutely without reference.

My position about theory of names is quite close to Frank Jackson’s as follows:

...What we do with them [viz. sentences containing proper names] makes it clear that we—we, the folk—know perfectly well that tokens of “*N* is *F*” stand at the information—delivering end of an information—preserving causal chain, sustained by the way our language community uses the token name “*N*” that figures in the sentence, a chain which starts with some kind of baptism of the object the information is about. The token name ties the sentence to the object the sentence gives information about via the causal chain. (Jackson 2010: 138)

Jackson once said: “...there have always been defenders of the description theory, and many of the things I say have been said in one form or another, somewhere or other, by someone or other” (1998: 201). His saying is almost completely applicable to my situation. When developing my SHCD, I have got different kinds of inspiration from different scholars, whether they are descriptivists or referentialists, e.g., Frege (1892), Russell (1905), Strawson (1950), Searle (1958, 1983), Donnellan (1970), Kripke (1972/1980), Dummett (1973, 1981), Evans

¹ The word “meaning” has a wide sense and a narrow sense in modern philosophy of language. In its wide sense, “meaning” includes both the sense [*Sinn*] and reference [*Bedeutung*] of a linguistic expression; in its narrow sense, “meaning” only denotes the sense of an expression being understood and grasped by human minds. This paper uses the narrow sense of “meaning”.

(1973), Putnam (1975), Plantinga (1978), Burge (1979), Devitt (1980), Lewis (1984), Kroon (1987, 2009), Stanley (1997), Jackson (1998, 2010), Devitt & Sterelny (1999), Soames (1998, 2002), Sosa (2001), Salmon (1986, 2005), Braun (2006), and so on. However, I want to emphasize addition that in my SHCD, I do not only choose something from what other scholars have said and combine them into an unitary theory, but I also make my own contribution. All of these will be clarified in what follows.

A1. *Names, including proper and general names, come from generalized “initial baptisms” of objects. We usually dub a physical object with a proper name by ostension, and dub a theoretical entity with a (general) name like “quark” by description.*

In *Naming and Necessity* (1980:71), Kripke reformulates six theses of cluster version of descriptivism as the target of his attack, and then states thesis (C):

For a successful theory, the account must not be circular. The properties which are used in the vote must not themselves involve the notion of reference in such a way that it is ultimately impossible to eliminate.

He explains further, “(C) is not a thesis but a condition on the satisfaction of the other theses” (1980: 71). Consider some examples clearly violating the noncircularity condition. Someone uses the name “Socrates”. How are we supposed to know to whom he refers? By using the description which gives the sense of it. According to Kneale (1962), the description is “the man called ‘Socrates’”. But this description tells us nothing at all. Taking it in this way it seems to be no theory of reference at all. We ask, “To whom does he refer by ‘Socrates’?” And then the answer is given, “Well, he refers to the man to whom he refers”. If this were all there was to the meaning of a proper name, then no reference would get off the ground at all (cf. Kripke 1980: 70).

Just as Kripke says, some descriptivists indeed make circular explanations in order to avoid the difficulties of finding appropriate description(s) which uniquely determine the referent of a proper name. For instances, metalinguistic descriptivism claims that the name “N” can be characterized by such descriptions as “the object called ‘N’”, “the bearer of ‘N’”, “the thing which is the bearer of ‘N’”, or “the self-same thing which is the bearer of ‘N’” (cf. Bach 1981: 372; Katz 1990: 40, 46), etc. Some causal descriptivists think that the referent of N is determined by some descriptions like “the individual which has been named ‘N’ in its initial baptism and whose name ‘N’ has been got handed down the causal chain” or “the object referred to by others in my linguistic community or by my interlocutors as ‘N’”. Clearly, these are the cases of reference-borrowing. Moreover, in order to determine the referent of “Aristotle”, descriptivists usually appeal to some description like “the teacher of Alexander”. The problem is how to determine the referent of the new name “Alexander” involved in the new description. If appealing to some description like “the most powerful one of Aristotle’s stu-

dents”, we obviously commit the fallacy of circular account. If we assert to other descriptions which possibly involve other names, the question of “how to determine the referents of other names?” will still come out. The former is a obvious circle, and the latter is an infinite regress.

Other descriptivists design descriptions such as “the entity that this body of information is about” (Forbes 1990: 538–539), “the subject of this mental dossier” (Nelson 2002: 415). They think that in order to determine the referent of N, at first we have to identify a body of information or a dossier (how to do this if we don’t know to which object N refers?), and then associate the body of information or the mental dossier with an object, and finally identify the object satisfying the body of information or the dossier as the referent of N. I think, there are too many, in Russell’s term, “zigzags”. Let’s consider two possibilities: (i) at the beginning, we collect information about an object, and then dub the object with name N, finally appeal to the information to determine the referent of N; (ii) at the beginning, we dub an object with N, later on collect the information about the object, finally other people appeal to the information to identify the referent of N. I’d like to ask a question: which of the two alternations is nearly right? My answer: (ii) is more close to be correct than (i).

Strawson says: “...one reference may borrow its credentials, as a genuinely identifying reference, from another; and that from another. But this regress is not [should not be] infinite” (1959: 182n). Searle also considers the parasitic use of names, that is, one speaker’s use of a name is parasitic on other speakers’ prior use of the name. He points out that the parasitic use of a name is not enough for determining the referents of the name, and that it must terminate in somewhere in order to determine which object the name designates (1983: 243–244). I agree, and I directly assert that reference-borrowing must terminate in the generalized initial baptisms of objects. Be a descriptivist or referentialist, all of us actually have the same starting-point: we dub an object with a name in the baptism of that object. After that, we differ in replying the following question: How do those people being absent from the dubbing event or the subsequent users of the name identify the referent of the name?

In the initial baptisms of objects, we dub a physical object in front of us with a name by ostension, or dub an unseen object with a name by description. As one example of the second way of naming, Le Verrier used descriptions to name an astronomical object, i.e. Neptune, which had not been found at that time. Without initial baptisms, descriptivists have no way to avoid circularity. It is Kripke’s thesis (C) that makes me, a firm (or stubborn?) descriptivist, be aware of this fact. Moreover, I want to emphasize that only after an initial baptism by which an object is named, does the object enter into our language and cognition. In most cases, we can’t talk about an object beyond our horizon without a name: what attributes does the object have? What similarity and difference

between one object and another are there? Thus, for people who did not participate in the baptism of an object, it is absolutely necessary for them to know something about the object to which a name refers. Since a very small amount of people participate in the dubbing event of a particular object—only for them dubbing determines reference, most people are not in position to identify the referent of a specific name by pointing; they have to appeal to descriptions to identify that object. So, it is of general interest and great significance to investigate the question of how names refer to their referents by means of descriptions.

Here, my position is quite close to Evans' in his (1982). In terms of the roles that participants in a name-using practice might play, Evans distinguishes what he calls “producers” from “consumers” in the practice of using name α to designate object o . A “producer” in the practice of using α to refer to o is somebody who “know o as α ”. S know o as α if and only if S has a specific kind of rapport with o , where the use of α forms part of this rapport, e.g. S has the capacity to identify o demonstratively and recognize o after breaks in observation. A “consumer” with respect to the practice of using α to refer to o is a participant in the practice who does not know o as α . “...it is reasonable to attribute to a speaker the intention to participate, by his use of a name, in the same practice as was being participated in by those speakers from whose use of the name the information he has associated with the name derives” (1982: 387).

Here, I have other two comments about descriptivism and naming.

First, if not considering the initial baptism of an object in which the object was dubbed with a name, in order to avoid the fallacy of circularity, descriptivists have to assert that a term has the referent it does just because it is associated with a set of descriptions in purely general, non-indexical or particular involving terms; these descriptions are uniquely satisfied by an entity, which then counts as the reference of that term. As Strawson argues, this is an impossible task for descriptivists to accomplish: an identifying description “need not be framed in purely general terms. In general, indeed, it could not be so framed; it is impossible, in general, to free all identification of particulars from all dependence upon demonstratively indicatable features of the situation of reference” (1959: 182n). If baptizing a object with a name was introduced into descriptivist picture, then, the participants of the baptizing know the referent of the name. Other people can borrow reference from the participants, that is, their use of the name is parasitic to the use by those participants. Furthermore, when they describe other objects to other people, they can use those names which they already know what they refer to and how they refer to their referents, more straightforwardly, they can use descriptions containing other names. As Stanley presents, at least for some descriptivists, “the descriptions which fix referents can, and indeed often must, contain non-descriptive elements” (1997: 564).

Second, naming an object is a social event. Not everyone can give a name to an object; instead, denominators must have appropriate social status. For example, only parents or respectful persons invited by the parents can dub a newborn baby with a name. Naming is also a public event, a “game” in which the object to be named, the denominators, and the witnesses are involved. Moreover, the spread of a name is also a social process. It should also be socially evaluated whether or not a name is appropriate to an object. If a name is not considered to be proper, the relevant object may be re-named. A person can have his “nickname” or “penname”. Sometimes the nickname or penname becomes so popular that the original was forgotten or only known by a small amount of people. Consider the names “Mark Twin” and “Samuel Clemens”. So to speak, the naming relation of a name to an object is socially conventionalized: it is not only semantic relations of names to objects, but also social relations among names, the corresponding objects, and our linguistic community. Generally speaking, there are three ways to guarantee that an object is successfully named: (i) People who have appropriate social status name an object by pointing; (ii) People who have appropriate social status name an object by description; (iii) Experts in their professional fields name a theoretic entity like quark with which ordinary people are not familiar.

A2. In the causal-historical chain of communication, the descriptive information about what name N refers to is passed on from one person to another and from one generation to the next; thus, the causal-historical chain of N is a chain starting from the dubbing of an object with N and preserving information about that object as N's bearer.

Kripke claims that after the initial baptism, “through various sorts of talk the name is spread from link to link as if by a chain” (1980: 91). He argues, the chain firstly and mainly transmits the referent of name N, although it also could transmit information about N’s bearer so that it could be a chain for transmitting information. When hearing N from somewhere, even though speakers at the far end of the chain are non-informed (ignorance), mis-informed (error), or poorly-informed (insufficiency) about the referent of N, they still can use N to refer to that object. Kripke states the condition as follows for successful transmission of reference in the chain, and accept the possibility that something is mistakenly transferred in the chain:

...When the name is ‘passed from link to link’, *the receiver of the name must, I think, intend when he learns it to use it with the same reference as the man from whom he heard it.* If I heard the name ‘Napoleon’ and decide it would be a nice name for my pet aardvark, I do not satisfy this condition. (Kripke 1980: 96; emphasis added)

...Obviously, the name is passed on from link to link. But of course not every sort of causal chain reaching from me to a certain man will do for me to make a reference. There may be a causal chain from our use of the term ‘Santa Claus’ to a certain historical saint, but still the children, when they use this, by this time probably do not refer to that saint. (Kripke 1980: 93)

I have three comments about the causal chain of N.

(1) The causal chain of N is a chain preserving information about N's bearer; only by means of preserving information about N's bearer, can the chain transmit the referent of N. Without the least information such as "N is an X" (here X is a sortal), e.g. "Dan is my pet dog", nobody can take any word s/he heard to be a name. Only based on such information can a hearer judge that the word s/he heard is a name of an object. The following citation from Kripke is puzzling to me:

...A mathematician's wife overhears her husband muttering the name 'Nancy'. She wonders, whether Nancy, the thing to which her husband referred, is a woman or a Lie group. Why isn't her use of 'Nancy' a case of naming? If it isn't, the reason is not indefiniteness of her reference. (Kripke 1980: 116n)

That is the alleged case of ignorance against descriptivism: even if one person know nothing about what thing a name designates, s/he still can use the name to refer to what it designates. Back to Kripke's example: although the mathematician's wife has no idea of what thing Nancy is, or of whom Nancy is, she still can use "Nancy" as a name to refer something or somebody. But I have serious doubt with this claim: how does the wife know that "nancy" is a name rather than a noise from her husband, since he also mutters something like "haha" and "bala"? why is "nancy" a name but "haha" or "bala" not? In my understanding, she takes "Nancy" as a name but does not take "haha" and "bala" as names, just because as one competent English speaker she has common sense that in English "Nancy" is usually used as the name of a female, but in few time "haha" and "bala" are used as the names of objects. However, "usually" doesn't mean "always", and "few" doesn't mean "never". Consider such a possibility: the mathematician pronounces "nancy" just for fun, exactly like he pronounces "haha" and "bala". All these "noises" serve the same purpose: to amuse himself and make himself relax. It is reported that some African people have very long and strange names: some words, such as "pain", "nuisance", "Good by", "Friday", are used in the names of African people. I think, if an agent is completely ignorant of a language, he has no reason to identify any word of that language s/he heard as the name of some object; also, if he has no information about an object to which a name refers and also cannot identify the referent of the name demonstratively, he has no reason to regard any word s/he heard as the name of that object, unless he names the object to which he faces by himself.²

Perhaps we should consider the cases of "Cicero" and "Feynman" discussed by Kripke (1980: 81). About what the name "Cicero" designates, many people know only that he was "a famous orator of ancient

² Evans also investigates what conditions have to be satisfied by an expression x and an item y for x to be a name of y. In his view, "NN" is a name of x if (and only if): (i) There is a community in which people use "NN" to refer to x; (ii) It is common knowledge that "NN" is so used; (iii) The reference in (i) relies on the knowledge in (ii), and not on the knowledge that x satisfies some predicate embedded in "NN" (cf. Evans 1973: 1, 18).

Rome”, and about “Feynman” only that he was “a physicist or something”. Obviously that such description(s) is not sufficient for fixing the referent of the corresponding name uniquely, but people still use it as the name of a person. My reply: Yes, people still use “Cicero” and “Feynman” separately as a name for *Cicero* and *Feynman*, because they have known that Cicero is a famous Roman orator, and Feynman is a physicist. Based on such little information, they know that both “Cicero” and “Feynman” are used by other people as the names of two human beings. Since the information is so poor and insufficient, they are not in position to pick out two men to which two names refer. However, as the members of their linguistic community, their uses of two names are parasitic to the uses by other members of the community. Although they don’t know exactly what individuals to which two names refer, but some other members know. They borrow reference from the other members of the community who know.

(2) In the causal chain of N, only information about N’s bearer can guarantee that the intention of present speakers is in accord with that of previous speakers.

In order to guarantee that the referent of a name is transferred successfully down the chain, Kripke just mentions one condition that in using names which s/he heard from other speakers a hearer must keep the same intention with the speakers’. Kripke stops here and does not make further enquiry. But I want to ask a question: how do we make sure that the condition will be satisfied? In my view, if N’s bearer is absent in the place of utterance, then speakers cannot transmit the referent of N to hearers by pointing. They have to say something about N’s bearer in order to make hearers know that they are talking about the object to which name N refers. Just as my analysis of the case of “Nancy” shows, if without the least necessary information, there will be no successful reference-transmission, even no name transmitted. This is the first point which I want to emphasize here.

Secondly, different information will result in different reference, even result in different names. For instance, two guys talk about a man named “John Lycan”, but one talks about a man born in a wealthy family, who himself is a distinguished professor of a well-known American university, publishes several good books, and often travels abroad to deliver lecture; another talks about a man born in an impoverished family, who himself is fortunately a gifted football player, makes a huge amount of money, and lives a quite decent life. Under such circumstance, two talkers will know soon that they are talking about different persons happened to have the literal “same” name(s).

Thirdly, mistaken or insufficient information will produce mistaken reference, called “reference-shift”. For instance, Evans mentions the case of “Madagascar”. Originally, it named a portion of the African mainland. But, *misunderstood* by Marco Polo, it became attached instead to the great island off the coast of Africa (Evans 1973: 11). Despite

the fact that there is a continuous “chain” of derived uses of the name ‘Madagascar’ going back to the baptism of the mainland, the name as used now refers to an island. The reason why the reference-shift happen is that the information is wrongly transmitted by Marco Polo. Kripke himself also mentions that “Santa Claus” might originally designate a certain historical saint, but today children use it to refer to a fictional figure in religion (Kripke 1980: 93). This is a case of “reference-failure”: a name from “referring” to “empty”, viz. not referring. Why does this phenomenon happen? One reasonable explanation is that there is no sufficient information transferred down to children today.

Fourthly, new information will result in new reference, even new names. I call such situation “reference-regeneration”. For instances, a certain name was originally taken to designate a mythological figure, but new archaeological evidence shows that the name refers to a real historical figure, so the name changes from “empty” to “referring”; Or a certain name originally designated a fictional figure, but later people used this term to refer to a real person, and this man was so famous in history that people forgot the fact that this name once denoted a fictional character. Certainly, in the cases of so-called “reference-regeneration”, actually there are two pairs of names which refer to two pairs of people, but we cannot ignore the fact that each pair of names is literally “same”, and there is some kind of continuous history in that pair.

I’d like to include reference-shift, reference-failure, reference-regeneration together under the title “reference-shift”. In my view, the fundamental reason why reference-shift happens is that when information about N’s bearer is transmitted down a causal chain of N, people commit some mistakes about the information of N’s bearer consciously or unconsciously.

Here, I want to talk something more about the causal chain of N. Actually, whether be descriptivists, such as Evans and Searle, or be referentialists, such as Kripke and Donnellan, there are quite many similarities between their conceptions of names: they both (at least *could*) agree that there are causal, historical, chains of communication, in which names get handed down from one person to another, from one generation to the next; and they both require intentional components (the intention to refer). What distinguish descriptivists from referentialists are their different answers to some key questions, e.g., *what is it that is transmitting down the causal chain about a name?* Clearly, it is not just the name; it is the name plus something else that is conventionally associated with the name. For descriptivists, what is conventionally associated with the name is a *sense* (or description, or cluster of descriptions, or way of picking something out); For Kripke, what is conventionally associated with the name is an *object*. Another key question is: *how are names connected to their referents?* Frege claims that there is an intermediary, i.e. a *sense*; Searle asserts that “objects are not given to us prior to our system of representation”, and so our

representations intervene between name and referent (Searle 1983: 326). But Kripke maintains that the connection is unmediated: names are directly referential. It is these points which distinguish descriptivists and referentialists apart.

A3. Only informative descriptions of N's bearer acknowledged by our linguistic community constitute the meaning or partial meaning of N. These descriptions describe the features of that object, and the collection is always open-ended and vague to some extent.

At first, I want to make clear what is really the semantic reference of a linguistic expression. I agree with Strawson's claim: "Mentioning', 'referring' is not something an expression does; *it is something that someone can use an expression to do*" (1950: 326; emphasis added). Even Kripke himself thinks that the semantic reference of a designator in a given idiolect (which usually includes a large linguistic community) is the thing which is determined by the conventions or rules of the idiolect together with facts about the world (e.g. which satisfies the descriptive property in question) on the occasion of use of the designator (cf. Kripke 1977: 111). In my view, since the conventions or rules of a language are the business of a linguistic community, the semantic referent of a designator could be said to be the thing to which *the linguistic community* takes the designator refer. Especially for a definite description, *its semantic reference is not the factual satisfier of that description, but the object which our linguistic community think satisfies that description*. I will make this idea clear further in what follows.

In my view, in the causal chain of name N, not all the informative descriptions of N's bearer are preserved: some are thrown away or forgotten, because they are not accepted as true by our linguistic community; some are revised, because they are partly true and partly false. Only those informative descriptions acknowledged by our linguistic community are preserved, getting handed down the causal chain. Finally, they become a part of public beliefs about that object, and enter into dictionaries or encyclopedias. In some sense, dictionaries or encyclopedias are just the summarization or refinement of our previous cognitive achievements, so they have experiential origins and contents, and can be enlarged, revised, or even replaced by our new epistemic achievements (cf. Chen Bo 2015: 103–104). It is important for agents to know the informative descriptions of N's bearer accepted as true by our linguistic community, because only these descriptions determine the referent of N, constitute a linguistic or cultural tradition about the use of N, and even a part of the capacity of a competent language-user to properly use N.³

³ Evans admits that there are indeed causal relations or causal chains with respect to the use of names, but Kripke "has mislocated the causal relation; the important causal relation lies between that item's states and doings and the speaker's body of information—not between the item's being dubbed with a name and the speaker's contemporary use of it" (Evans 1973: 13).

I will introduce symbols to characterize the meaning of name N: let lowercase letters, $a, b, c, d, e, f, g, h, i, j, k, \dots$ separately stand for a description of N's bearer. Some descriptions are not accepted as true by our linguistic community, so they will not enter into the collection of descriptions as the meaning of N; only those description acknowledged by our linguistic community enter into the collection about N's bearer: $\{a, b, c, d, e, f, \dots\}$, in which “...” shows that there are some members outside of the listed, and we can change the members of the collection if necessary, that is, let some old member(s) get out, and some new member(s) come in, if we get new evidence; so the collection is always open-ended and is vague to some extent. Since the collection illustrates the consensus of our linguistic community about N's bearer, and generally acknowledged by our linguistic community, so an operator * for consensus can be put in the front of the collection as a superscript: $*\{a, b, c, d, e, f, \dots\}$. This kind of collection of descriptions determines the referent of N. Of course, we could have some other collections of descriptions of N's bearer by means of our counterfactual imagination, e.g., $\{-a, -b, -c, -d, -e, f, g, h, j, k, \dots\}$, $\{-a, b, -c, -d, e, -f, u, v, w, x, \dots\}$, in which ‘-a’ shows that a is absent, and so forth. These collections do not constitute the meaning or partial meaning of N, because they have not been agreed by our linguistic community. We can't use them to determine the referent of N, at least we can't use them to identify the object to which we *usually* use N to refer.

From this perspective, the so-called “counterexamples”, such as Gödel/Schmidt case, Peano/Dedekind case, Johna-Moses-Aristotle cases, of descriptivism given by Kripke in his semantic argument, could be explained away. Here, I will consider the first two.

Kripke conceives a counterfactual situation. Gödel had a friend called “Schmidt”, who had actually proved the incompleteness of arithmetic. But Gödel somehow got hold of the manuscript and published it in his own name. Then Gödel achieved fame as “the man who discovered the incompleteness of arithmetic”. However, in fact, the real referent of that description is the man Schmidt. If “Gödel” is synonymous with the description “the man who discovered the incompleteness of arithmetic”, does “Gödel” change its referent into the man Schmidt? Kripke replies “No”, “Gödel” still designates the person called “Gödel” whereas the description “the man who discovered the incompleteness of arithmetic” refers to the man *Schmidt*, because Schmidt is *actually* the person satisfying that description, and we make a mistake when using the description to refer to Gödel.

I can reply to Kripke as follows. In your argument, I find an implicit supposition: the question of “how does some description(s) refer to an object?” just concerns the relation between the description(s) and its satisfier, between a language and the world, which are only the matters of fact, and has nothing to do with the intentions, conventions and customs of our linguistic community in using the description(s) and the

language. In other words, the semantic referent of some description(s) is just the object which *in fact* satisfies the description(s), rather than the object to which *our linguistic community takes the description(s) to refer*. But I have argued that this supposition is wrong (cf. Chen Bo 2013a: 423–433). Here, I just reply Kripke very shortly: your fabricated story is not acknowledged by our linguistic community; your fancy about Gödel is not in the causal chain of the name “Gödel”. Therefore, we can still believe that the description “the man who discovered the incompleteness of arithmetic” designates the man *Gödel* rather than the man *Schmidt*. However, if your imagined situation is supported by good evidence and agreed by our linguistic community, perhaps we will cut off the connection of the name “Gödel” with the description “the man who discovered the incompleteness of arithmetic”, and establish a new connection of the description with the name “Schmidt”. Perhaps we will also build up the connection of the name “Gödel” with the new description “the notorious man who stole Schmidt’s proof of the incompleteness of arithmetic”. Just as Kripke himself says,

In general our reference depends not just on what we think ourselves, but on other people in the community, the history of how the name reached one, and things like that. It is by following such a history that one gets to the reference. (Kripke 1980: 95)

Kripke also talks about Peano-Dedekind case. It is commonly believed that Peano is the man who discovered certain axioms which characterize the sequence of natural numbers. But actually it is Dedekind who discovered these axioms earlier; thus the description “the man who discovered certain axioms which characterize the sequence of natural numbers” denotes Dedekind. Many people mistake Einstein for both the discoverer of the theory of relativity and the inventor of the atomic bomb. But actually it was not a single person but a group of people who invented the atomic bomb. Similarly, many people regard Columbus as the first man to know that the earth was round and the first man to discover America. However, there might have been someone else who is the semantic referent of these descriptions, whereas “Columbus” still refers to the person originally called “Columbus”.

My reply is similar to the Gödel-Schmidt case. What is of great significance is not what Peano, Einstein, and Columbus have actually done, but what is acknowledged by our linguistic community. Our community even makes a series of institutional arrangement about academic acknowledge, such as anonymous referee, open-access publication, objections and replies, discussion and debate, citation data as influential factor, public reward system, and so on. Only those descriptions of a scientist agreed by our community can be regarded as the part of the “official” history of the person and constitute the meaning or partial meaning of the relevant name. In contrast, those descriptions of the person rejected by our community will be forgotten, or just become

the topics of chat, gossip, or casual conversation at leisure time. We never consider those descriptions seriously.

In sum, my idea is this: the meaning of a name depends on consensus of our community rather than somebody's wild imagination, and a name or description designates what our language community agrees to use it to designate. In semantics, there is no pure matters of fact, the intentionality of a linguistic community must be considered.

A4. *With respect to the practical needs of agents, there is a weighted order in the collection of informative descriptions of N's bearer; that is, in the collection some descriptions are more important or central than others for determining the referent of N.*

When restating the cluster version of descriptivism about names, Kripke mentions thesis (3): "if most, or a *weighted* most, of the ϕ 's are satisfied by one unique object y , then y is the referent of 'X'" (Kripke 1980: 71, emphasis added). That is to say, traditional descriptivists do not give equal weight to all the descriptions in the collection. When identifying what a name designates, some descriptions are more important or central than others. Evans thinks, the denotation of a name is fixed by the bodies of information; a particular object is the *dominant source* of the descriptions we associate with the name, and it is the *dominant* description that plays a *crucial role* in determining the referent of a name (Evans 1973: 15–17). Putnam notices that the stereotype of "tiger" includes such features as "being an animal", "being big-cat-like", "having black stripes on a yellow ground". He assumes that the feature "being an animal" is more central than others, because it is impossible to conceive that tigers might not have been animals (Putnam 1975: 188–190). In my view, the members of the collection of descriptions about N's bearer have to be organized into some kind of structure.

I suspect that Kripke might implicitly hold a similar position. For him, most descriptions are non-rigid designator because they usually describe the superficial or accidental features of their objects; however, some descriptions are rigid designator, such as "the positive odd number less than 2", "the element with the atomic number 79" and "H₂O", because they characterize the essence of the corresponding objects. Essence is what an object or natural kind necessarily has, i.e. what it has in all possible worlds in which it exists. According to Kripke, the essence of an individual such as "Aristotle" is its origin; the essence of an artefact such as "table" is its constituent material; and the essence of a natural kind like "tiger" is its internal structure. The descriptions about essence will refer to an object or a natural kind in all possible worlds in which the object or the kind exists, so they are rigid designators. Thus, essential descriptions of an object are more weighted than others.

I myself also think that not all the descriptions in the collection have equal weight. When determining the referent of a name, some descriptions are more important than others because they are essential descriptions. In a long interview by BBC, Quine says that so-called

“essence” is what is most important; the essence of a thing depends on how the thing is described. But he thinks that since we could not make it clear that what is the most important about an object, we could not explain clearly what essence is. So he rejects essentialism, taken it as a notorious form of Platonism. Putnam says that “importance is an interest-relative notion” (Putnam 1975: 157), that is, it depends on our interest to decide which properties more important than others. I applaud to this brand of essentialism. I think that the importance is relative to the agent’s interest. By introducing the parameter “with respect to humans’ interest”, I will relativize and thus diversify the essence of an object. For example, the essence of human beings for zoologists is different from that for sociologists. If we can generalize common need of humans’ cognition and practice, then we could find out the general essence of an object. For instance, the general essence of human beings might be “the animals that are able to speak, to think, and to make tools”. This version of essentialism can be called “interest-relative essentialism”, whose details and arguments have to be left to other papers.

In his paper (2011), Costa criticizes the traditional cluster theory of proper names because a cluster has no internal structure, being completely disordered: all descriptions belonging to the cluster seem to have the same value and play the same identifying role. He wants to put an order into this mess. He distinguishes all the descriptions of an object in the cluster into two groups. One consists of fundamental descriptions, including “(i) a localizing description, which gives the spatio-temporal location and career of the object, and (ii) a characterizing description, which gives what are considered the most relevant properties of the object, those that give us the reason to use the name in referring to it” (Costa 2011: 260). Another consists of auxiliary descriptions which seem to connect a name with its bearer in a more or less accidental fashion. The second group includes metaphorical descriptions, accidental but well-known descriptions, accidental and usually unknown descriptions, adventitious descriptions (Costa 2011: 261–262). Then, Costa formulates a meta-descriptive rule MDR:

A proper name *N* is used to refer to the object *x* belonging to a certain class *C* of objects, *iff* it can be assumed that *x* properly originates our awareness that

- (i–a) *x* satisfies its localizing description for *N*, and/or
- (i–b) *x* satisfies its characterizing description for *N*, and
- (ii) *x* satisfies the description(s) sufficiently, and
- (iii) *x* satisfies the description(s) better than any other object belonging to *C*.

(Costa 2011: 270)

Obviously, Costa and I have a similar viewpoint that there is a structure in the descriptions-cluster of an object. But my position is quite different from his in other sides, e.g. I pay much more attention to the

role of a linguistic community for determining the meaning and reference of a name than he does. In my view, first, the distinction of fundamental and auxiliary descriptions makes sense only with respect to the practical needs of agents. Take “Aristotle” for an example. We can characterize Aristotle as a famous scholar, a philosopher, a linguist, a biologist, an educator, ... so the relevant characterizing and auxiliary descriptions about him will be radically different: fundamental descriptions in one encyclopedia will become auxiliary ones in another, vice versa. Secondly, only those descriptions acknowledged by our linguistic community can become the meaning or partial meaning of the name which refers to the object. Thirdly, we cannot exactly determine how many descriptions an object has to satisfy in order to be the referent of a name, since in replying this issue, we consider not only the quantity of descriptions, but also the order of descriptions in the cluster, and even the practical needs of agents.

A5. The meaning or even partial meaning of N, together with the background of a discourse, the network of knowledge, speaker’s intention, etc., determines the referent of N.

Kripke puts forward his semantic argument against descriptivism as follows:

- P1 If descriptivism is correct, then, the meaning of name N, which is given by one description or a cluster of descriptions, should provide a set of necessary and sufficient conditions for determining what N designates.
- P2 In fact, the corresponding description(s) cannot supply such a set of conditions for fixing the referent of N.
- C Descriptivism is wrong.⁴

I judge that P1 of the semantic argument of Kripke’s does not hold, because it relies on a problematic assumption, namely, descriptivists have to hold two claims: (i) If name N has its meaning and the meaning is given by some description(s), the description(s) should provide a set of necessary and sufficient conditions for determining the referent of N; (ii) It is possible for us to find out such a set of necessary and sufficient conditions for determining N’s bearer. I have argued that the assumption is wrong (cf. Chen Bo 2013a: 435–438).

In my view, when determining what a name designates, it is absolutely necessary to fix different domains of discourse in different contexts. These domains are usually smaller than the Universe containing all the actual individuals in the external world, and much smaller than the Super-Domain containing all the possible individuals in all possible worlds. When determining the referent of a name by means of its meaning, actually we choose the referent from the specific domain of

⁴ Salmon regards the semantic arguments as “the strongest and most persuasive of the three kinds of arguments for the primary thesis of the direct reference theory” (Salmon 2005: 29).

discourse rather than always from the Universe or the Super-Domain. Under such contexts of utterance, only a few ordinary descriptions are required to identify the referent of a name, since only a finite number of individuals are in that place.

For example, “the girl dressed in red clothes” is not enough to determine the referent of any name, since there are too many girls dressed in red clothes in the world, let alone the amount in all possible worlds. However, there are only a small number of people in a particular context. When one asks “who is Lori?” someone else replies that “Lori is the girl dressed in red clothes”. If there is exactly one girl dressed in red clothes in that place, only by using the description about surface feature of a person, can we identify to whom the name “Lori” refers. If there happen to be many girls dressed in red clothes, we can keep talking to give more descriptive information in order to identify the referent of “Lori”.

I agree with Searle’s idea that speakers’ intention, Network and Background play a crucial role in determining what a name designates. Network includes personal convictions, scientific knowledge, and the existence of social practices and institutions, and it is in virtue of the network that humans succeed in having meaningful experiences or saying meaningful things. Background is the set of abilities, capacities, tendencies, and dispositions that humans have; it itself is non-representational and non-intentional. For example, when someone invites me to attend his/her wedding, I know that I have to dress formally and bring him/her significant gift(s); when someone invites me to join a country music, I know that I can dress casually and behave quite wild, even though the obvious request does not include this kind of details. Background beliefs give clues to my judgment and choice. Here, just consider one example as follows.

Donnellan (1970: 335–58) makes a bold envisagement. Suppose that all that a certain speaker knows or thinks he knows about Thales is that he is the Greek philosopher who said that all is water. But suppose that there was no Greek philosopher who said such a thing, and Aristotle and Herodotus were referring to a well digger who said, “I wish all was water so I wouldn’t have to dig these damned wells”. Further, suppose that there was a hermit who had no contact with anyone, who actually held that all was water. Furthermore, suppose that Herodotus had heard a frog at the bottom of a well making croaking noises that sounded like the Greek for “all is water”; this frog happened to be a family pet named “Thales”, and this incident is the origin of the view that somebody held that all is water. Then, we will meet a serious question: when using the name “Thales”, do we refer to the Greek philosopher, the well digger, the hermit, or the frog? Searle argues that in order to answer this question, we have to rely on the relevant Network of Intentionality. When we say “Thales is the Greek philosopher who held that all is water”, we do not just mean anybody who held that all is water, we mean *that person* who was known to other Greek philoso-

phers as arguing that all is water, who was called in his time or subsequently by people as “Thales”, whose works and ideas have come down to us posthumously through the writings of other authors, and so on.

...in all these cases there will be an external causal account of how we got that information, but what secures reference is not the external causal chain, but the sequence of the transfer of Intentional contents. The reason we are not tempted to allow the hermit to qualify as Thales is that he simply does not fit into the Network and the Background. (Searle 1983: 252–253)

I think that in a specific context, sometimes we can determine what a name designates just by one description, while sometimes we can achieve this by a cluster of descriptions. Can we generally explain how many descriptions we need to determine the referent of a name? No, because we also have to consider the speaker’s intention, Network and Background when determining the referent of a name. Therefore, just like the question “how does a name designate an object?”, the question “how do we identify what a name refers to?” is also relative to many social factors; it depends on the interplay of these factors to determine the referent of a name.

I think, it is the right place to reply shortly Kripke’s epistemic argument against descriptivism. The argument runs like this:

- P1 If descriptivism is correct, that is, name N is synonymous with its relevant description “the F”, then “N is the F” should be knowable *a priori*.
- P2 In fact, “N is the F” is not knowable *a priori*.
- C Descriptivism is wrong.

For example, consider two sentences:

- (1) Aristotle is Aristotle.
- (2) Aristotle is the teacher of Alexander the Great.

Kripke thinks, according to descriptivism, “Aristotle” is synonymous with the description “the teacher of Alexander the Great”; then, if substituting the second occurrence of “Aristotle” with “the teacher of Alexander the Great” in (1), we get (2). Since (1) is knowable *a priori*, so is (2). Actually, (2) is essentially an empirical statement, we have to judge its truth value completely based on historical documents and other empirical evidence. So, (2) is absolutely not knowable *a priori*. Therefore, descriptivism is wrong.

I have two comments about the epistemic argument:

First, I don’t think descriptivists have to hold such position that a name is synonymous with some relevant description(s). I take myself as a firm descriptivist, but don’t accept the synonymy thesis that “Aristotle” is synonymous with “the teacher of Alexander the Great”. In my view, the meaning of name N consists in the collection of informative descriptions of N’s bearer acknowledged by our linguistic community, and these descriptions describe the features of the object. Since the object as the referent of N and our cognition about that object are always in the process of change, so the collection of informative descriptions of

N's bearer is open-ended, and is vague to one degree or another. Therefore, N cannot be strictly synonymous with any definite description, even with the collection of such descriptions.

Second, even if we temporarily accept the synonymy thesis, we still cannot get the conclusion that (2) is knowable *a priori*. As I argued in Chen Bo (2015: 106–108), *semantic knowledge is empirical and encyclopedic knowledge, including the uses of language accepted by a linguistic community; it is the condensation, refinement, and summarization of our previous epistemic achievements, so it has empirical content and origin*. Quine emphasizes: “The lexicographer is an empirical scientist, whose business is the recording of antecedent facts; and if he glosses ‘bachelor’ as ‘unmarried man’ it is because of his belief that there is a relation of synonymy between these forms, implicit in general or preferred usage prior to his own work” (Quine 1953: 24). Why we can substitute “Aristotle” with the description “the teacher of Alexander the Great” in (1)? Because empirical evidence shows us that Aristotle is the teacher of Alexander the Great, we make use of this empirical message to do the substitution, then we get (2). So, (2) is also based on empirical evidence, and is just knowable *a posteriori*. For more details, see Bo (2013b).

Let *a* is a proper name, *b* is the corresponding description relevant with *a*, I can generalize the form of Kripke's epistemic argument against descriptivism as follows:

(1) It is knowable *a priori* that *a* is *a*;

(2) *a* = *b*;

So, (3) It is knowable *a priori* that *a* is *b*.

Kripke argues, since it is not knowable *a priori* that *a* is *b*, we should deny the descriptivist premise (2); Therefore, descriptivism about name is wrong.

But in my judgement, this argument is not sound, because it makes use of the principle of substitution which is problematic: $KF(a) \wedge (a=b) \rightarrow KF(b)$, here “K” means “know”. Rather, It should appeal to the valid principle of substitution: $KF(a) \wedge K(a=b) \rightarrow KF(b)$. (3) can follow not from (1) and (2), but from (1) and (2'): It is knowable *a priori* that *a* is *b*. Since descriptivists don't accept (2') as true, Kripke's epistemic argument collapses.

A6. *All names have their own referents, including physical individuals, and parasitic, fictional, or intensional objects. So there are few names absolutely without reference.*

In my view, the referential relation between a name and an object is not an objective relation between the two; on the contrary, a complete understanding of the referential relation of a name and an object involves three elements: speakers' intention, the meaning of name N, and the object to which N refers. Which object N designates depends on what a speaker intends to use N to designate. Moreover, what names

designate can be classified as follows: physical objects, parasitic objects, fictional objects, and intensional objects. The last three groups may be called “abstract objects”.

Obviously, in our language, many names refer to physical objects which exist in the actual world, i.e. in space and time, can be perceived by us, and have causal effect with each other. For example, there are names of people, such as “Socrates” and “Einstein”; names of natural objects, such as “Sun” and “Earth”; names of places, such as “Oxford” and “Tokyo”; names of countries, such as “China” and “United States”; names of organizations or political parties, such as “UNESCO”, “Japan’s Liberal Democratic Party”; names of books, such as “*The Organon*” and “*Origin of Species by Means of Natural Selection*”; names of events, such as “American War of Independence” and “the Second World War”. And so on. Physical objects are very close to “primary substance” called by Aristotle: “Substance, in the truest and primary and most definite sense of the word, is that which is neither predicable of a subject nor present in a subject; for instance, the individual man or horse” (*Categories*, 2^a13–14). “Moreover, primary substances are most properly called substances in virtue of the fact that they are the entities which underlie everything else, and that everything else is either predicated of them or present in them” (*Categories*, 2^b14–17). Moreover, physical objects include theoretical entities in natural sciences, such as atoms, electrons, photons, and other particles, which cannot be directly perceived by humans, but can be discerned by means of instruments.

There are also names designating parasitic objects supervened on physical individuals. Individuals come first, but they are not bare particulars without any property or quality. An individual itself has certain properties and also is related to other individuals. Individuals can be classified into different kinds or classes, such as animal, human being, and plant. In biology, there is a classification system consisting of species, genus, family, order, class, phylum, and kingdom. There are natural kind terms such as “cat”, “tiger”, and “lion”. Without natural kind terms, we have to meet serious difficulties in our ordinary talk, and even our scientific system will collapse. Although kinds or classes are the results of abstract thinking, they still have some kinds of objective existence. An object has property, and there are some relations between or amongst objects; these constitute so-called “states of affairs” or “facts”. Although states of affairs or facts are different from individuals since they are very difficult to be individualized, they are still objective. What mass terms such as “gold”, “wood”, “water”, “fire”, and “soil” designate cannot be individualized either, but they certainly exist in the actual world. Physical individuals always exist in space and time. Moreover, everything is in process of change, and their change follows regularities and laws. Since things are objective, so are the regularities and laws followed by them. In this way, we have a variety of entities supervened on or being parasitic to physical objects, such as qualities,

relations, classes or kinds, laws, etc. Certainly, these entities do not exist in space and time as substances, but it is reasonable to affirm that there are such kinds of supervened or parasitic entities; otherwise, physical objects will become pure abstraction or nothingness. Besides, there are another kind of abstract objects, such as natural numbers, real numbers, and complex numbers.

There are names denoting fictional objects, which do not exist in the actual world, but are created by human intellects. For example, there are various characters in Greek mythology, such as Gaea, Zeus, Poseidon, Apollo, Athena, Hermes, Dionysus; various characters in science fiction such as Superman, Spiderman, Harry Potter, Batman; a variety of literary figures such as Hamlet and Sherlock Holmes. Names denoting such kinds of objects are usually called “empty names”, because the objects to which they refer are not real, i.e., not exist in the actual world. The phrase “empty names” may come from Russell’s “robust feeling for reality”:

Logic, I should maintain, must no more admit a unicorn than zoology can; for logic is concerned with the real world just as truly as zoology, though with its more abstract and general features. (Russell 1919: 169)

I think that Russell’s position conflicts with our linguistic intuition and common sense. In natural languages, there are many names denoting fictional characters. We usually consider that these names are referring to something rather than nothing, because we can talk and exchange our opinions about them understandably. Besides, some mythological and literary figures have played very important roles in shaping the cultural identity of a nation.⁵ Why does logic, philosophy, and semantics exclude these names? Is the talk about them beyond the limit of reason? I do not think so, I do not like the phrase “empty names”.

There are names denoting intensional objects, including concepts, propositions, beliefs, thoughts, theories, and doctrines, etc. For instances, the concept “prime number,” the proposition that no bachelor is married, Archimedes Principle, Law of Universal Gravitation, Social Contract Theory, and Pragmatism. These objects depend on our linguistic actions, and can be grasped by different human minds, so they are inter-subjective. There are fierce debates about the existence of such kind of objects. We often meet two extremes: one is held by Frege and Popper, admitting objective thoughts or knowledge as entities; and the other is by Quine, rejecting any intentional entities like meaning and proposition.

I call physical and parasitic objects “actual existence” or “reality”. There are causal relations among actual objects and between actual objects and human beings. We can give the following criteria for “reality”: all actual objects have causal effects on perceivable material bodies, and we explain the changes of these material bodies by means of such

⁵ In his paper “Nonexistence” (1998), Salmon acknowledges the existence of literary figures like Sherlock Holmes and mythical objects like Vulcan.

effects. For example, force, macro-objects like Earth and human beings, micro-objects including atoms and other basic particles, are related together and have mutual effects; thus we admit the reality of force and atoms. So, in my opinion, “the actual” include both concrete objects like physical individuals, and at least some parasitic objects, such as properties, relations, classes, and laws. Moreover, I call fictional and intensional objects “ideal objects”, existing in humans’ epistemic system articulated by language. Ideal objects can be shared by different people, and occur as the products of human intellect. In addition, there are delicate relations between ideal existence and actual existence. In some sense, ideal objects are the reconstruction of actual objects by cognitive subjects in a variety of ways. Even for the queerest and strangest creations of human thinking, we can still discern the shadows of actual objects on them. As Popper emphasizes, once ideal objects are created by people, they usually transcend their producers and get their own independent lives.

* * *

I support Russell’s view that a logical and semantic theory may be tested by “its capacity for dealing with puzzles” (Russell 1905: 484). So, in order to test the effectiveness of my SHCD, we can examine how it reply to Kripke’s arguments against descriptivism, including the epistemic, the semantic and the modal, and other logical puzzles about names, e.g. the puzzle about belief presented by Kripke (1979), and what differences there are between my SHCD and other versions of descriptivism, and between SHCD and referentialism in dealing with these matters. However, all these tasks are far beyond the space-limit of this paper, and also some of them have been done in my other published papers (Chen Bo 2011, 2012, 2013a, 2013b, 2015), at whom some reader, if interested, may have a look.

References

- Bach, K. 1981. “What’s in a Name?” *Australasian Journal of Philosophy* 59 (4): 371–386.
- Braun, D. 2006. “Names and Natural Kind Terms.” In Ernest Lepore and Barry Smith (eds.), *Handbook of Philosophy of Language*. Oxford and New York: Oxford University Press: 490–515.
- Burge, T. 1979. “Individualism and the Mental.” In P. French, T. Euhling and H. Wettstein (eds.), *Midwest Studies in Philosophy*. Minneapolis: University of Minnesota Press: 73–121.
- Chen Bo. 2011. “Proper Names, Contingency *A Priori* and Necessity *A Posteriori*.” *History and Philosophy of Logic* 32 (2): 119–138.
- . 2012. “Descriptivist Refutation of Kripke’s Modal Argument and of Soames’s Defense.” *Theoria: A Swedish Journal of Philosophy* 79 (3): 225–260.

- _____. 2013a. "Kripke's Semantic Argument against Descriptivism Reconsidered." *Croatian Journal of Philosophy* 13 (39): 421–445.
- _____. 2013b. "Kripke's Epistemic Argument against Descriptivism Revisited." *Journal of Chinese Philosophy* 40 (3–4): 486–504.
- _____. 2015. "Social Constructivism of Language and Meaning." *Croatian Journal of Philosophy* 15 (43): 87–113.
- Costa, C. F. 2011. "A Metadescriptivist Theory of Proper Names." *Ratio* 24 (3): 259–281.
- Devitt, M. 1981. *Designation*. New York: Columbia University Press.
- Devitt, M. & Sterelny, K. 1999. *Language and Reality*. Cambridge, MA: MIT Press, 2nd ed.
- Donnellan, K. 1970. "Proper Names and Identifying Descriptions." *Syntheses* 21 (3–4): 335–58.
- Dummett, M. 1973. *Frege: Philosophy of Language*. New York: Harper & Row; 2nd edition, 1981. The citation is from the second edition.
- _____. 1981. *The Interpretation of Frege's Philosophy*. London: Duckworth.
- Evans, G. 1973. "The Causal Theory of Names." *Aristotelian Society Supplementary* 47: 187–208; reprinted in his *Collected Papers*. Oxford: Clarendon Press, 1985: 1–24. The reference number is of the reprinted.
- _____. 1982. *The Varieties of Reference*. Ed. John McDowell. Oxford: Oxford University Press.
- Frege, G. 1892. "On *Sinn* and *Bedeutung*." Reprinted in Michael Beaney (ed.) *The Frege Reader*. Oxford: Blackwell Publishers, 1997: 151–171.
- Forbes, G. 1990. "The Indispensability of *Sinn*." *Philosophical Review* 99 (4): 535–63.
- Jackson, F. 1998. "Reference and Description Revisited." In J. E. Tomberlin (ed.), *Philosophical Perspectives 12: Language, Mind and Ontology*. Oxford: Blackwell.
- _____. 2010. *Language, Names, and Information*. West Sussex and Oxford: Wiley-Blackwell.
- Kaplan, D. 1973. "Bob and Carol and Ted and Alice." In Hintikka, et al. (eds.) *Approach to Natural Language*. Dordrecht and Boston: Reidel: 490–518.
- Katz, J. J. 1990. "Has the Description Theory of Names Been Refuted?" In G. Boolos (ed.), *Meaning and Method: Essays in Honour of Hilary Putnam*. Cambridge: Cambridge University Press: 31–61.
- Kneale, W. 1962. "Modality *de dicto* and *de re*." In E. Nagel, P. Suppes and A. Tarski (eds.), *Logic, Methodology and Philosophy of Science, Proceedings of the 1960 International Congress*. Stanford: Stanford University Press: 622–33.
- Kripke, S. 1972/1980. "Naming and Necessity." In D. Davidson and G. Harman (eds.), *Semantics of Natural Language*. Dordrecht: D. Reidel Publishing Co.: 253–355. *Naming and Necessity*. Cambridge, Mass.: Harvard University Press, 1980.
- _____. 1979. "A Puzzle About Belief." In A. Margalit (ed.), *Meaning and Use*. Dordrecht: Reidel: 239–283.
- Kroon, F. 1987. "Causal Descriptivism." *Australasian Journal of Philosophy* 65 (1): 1–17.

- . 2009. “Names, Plans, and Descriptions.” In David Braddon-Mitchell and Robert Nola (eds.), *Conceptual Analysis and Philosophical Naturalism*. Cambridge, Mass.: MIT Press: 139–158.
- Lewis, D. 1984. “Putnam’s Paradox.” *Australasian Journal of Philosophy* 62 (3): 221–36.
- Lowe, E.J. 2007. “Does the Descriptivist/Anti-descriptivist Debate Have Any Philosophical Significance?” *Philosophical Books* 48 (1): 27–33.
- Nelson, M. 2002. “Descriptivism Defended.” *Noûs* 36 (3): 408–436.
- Plantinga, A. 1978. “The Boethian Compromise.” *American Philosophical Quarterly* 15 (2): 129–138.
- Putnam, H. 1973. “Meaning and Reference.” *The Journal of Philosophy* 70 (19): 699–711.
- Quine, W. V. 1953. *From A Logical Point of View*. Cambridge, Mass.: Harvard University Press. Second edition, revised, 1961. New York: Harper Row, Publishers. The reference number is of 1961 version.
- Russell, B. 1905. “On Denoting.” *Mind* 14 (56): 479–493. Reprinted in his *Logic and Knowledge*. Ed. R. C. Marsh. London: George Allen and Unwin, 1956: 41–56.
- . 1911. “Knowledge by Acquaintance and by Description.” In his *Mysticism and Logic and Other Essays*. London: Longmans, Green, and Co: 209–232.
- . 1919. *Introduction to Mathematical Philosophy*. London: George Allen and Unwin.
- Salmon, N. 2005. *Reference and Essence*, 2nd edition. Amherst, New York: Prometheus Books.
- . 1986. *Frege’s Puzzle*. Cambridge, Mass.: MIT Press.
- . 1998. “Nonexistence.” *Noûs* 32 (3): 277–319.
- Searle, J. 1958. “Proper Names.” *Mind* 67 (266): 166–73.
- . 1983. “Proper Names and Intentionality.” In his *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press: 231–261.
- Soames, S. 1998. “The modal argument: wide scope and rigidified descriptions.” *Noûs* 32 (1): 1–22.
- . 2002. *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*. New York: Oxford University Press.
- Sosa, D. 2001. “Rigidity in the Scope of Russell’s Theory.” *Noûs* 35 (1): 1–38.
- Stanley, J. 1997. “Names and Rigid Designation.” In B. Hale and C. Wright (eds.), *A Companion to the Philosophy of Language*. Oxford: Blackwell Publishers: 555–585.
- Strawson, P.F. 1950. “On Referring.” *Mind* 59 (235): 320–344.
- . 1959. *Individuals: An Essay in Descriptive Metaphysics*. London: Methuen.

Becker, Ramsey, and Hi-world Semantics. Toward a Unified Account of Conditionals

CHENG-CHIH TSAI

*Center for Holistic Education, Mackay Medical College,
New Taipei City, Taiwan*

In Lowe (1995), instead of endorsing a Stalnaker / Lewis-style account of counterfactuals, E. J. Lowe claims that a variation of C. I. Lewis's strict implication alone captures the essence of everyday conditionals and avoids the paradoxes of strict implication. However, Lowe's approach fails to account for the validity of simple and straightforward arguments such as 'if $2=3$ then $2+1=3+1$ ', and Heylen & Horsten (2006) even claims that no variation of strict implication can successfully describe the logical behavior of natural language conditionals. By incorporating the German logician O. Becker's modal intuition with the insight of Ramsey's Test, we show that there does exist a unified, strict-conditional based account of everyday conditionals, which withstands all attacks previously raised against truth-conditional accounts of conditionals. Furthermore, a subtle distinction between autistic and realistic readings of the indexical 'I' involved in a conditional helps us resolve a recent debate concerning the Thomason conditionals.

Keywords: Becker's Semantics, Ramsey's Test, Hi-world Semantics, Moore's Principles, Thomason conditionals.

1. *Introduction*

In Lowe (1995: 57), E. J. Lowe reckons

Conditionals in general present an extremely perplexing set of linguistic phenomena which often seem to defy a simple, uniform treatment of them for logical purpose.

Nevertheless, what he actually did was to try the seemingly impossible, namely, to defend a relatively simple core theory for them. Heylen

& Horsten (2006) defied Lowe's attempt and proved in general that no future attempts along the line of variation of strict implication would ever succeed. I think Lowe's attempt was indeed problematic, but Heylen and Horsten's analysis was problematic as well, because it was misled by an unwarranted assumption concerning possible worlds, which we shall explain in more detail later.

In Lowe (1983), Lowe expresses his general uneasiness towards the possible-world based account of conditional that were developed in the works of Stalnaker and Lewis.¹ As he wrote at that time,

At no time, however, shall I argue for my position by appeal to considerations involving 'possible worlds', because I find this notion so fraught with epistemological and ontological difficulties that to explicate conditional in terms of possible worlds must in my view, be to explain the obscure by the still more obscure. (Lowe 1983: 358)

Stalnaker and Lewis do employ possible worlds in their accounts of conditionals, but it is possible that what makes Lowe uneasy about possible worlds is not that the notion of possible worlds in itself is problematic, but rather that in order to cope with the phenomenon of conditionals, Stalnaker and Lewis have resorted to some additional structures imposed upon possible worlds, such as "worlds closest to ours", and "constantly varying spheres of possible worlds", etc. This partly explains the fact that twelve years later, in Lowe (1995), Lowe himself adopts a possible-world interpretation for counterfactuals as well—apparently, what he finds unacceptable are some miscellaneous notions associated with possible worlds, rather than possible worlds themselves.

As is remarked in Copeland (2002), in the early days of possible worlds, a Beckerean notion of possible worlds—or case-classes (Becker 1952)—is a strong contender along with the familiar notion of Kripkean possible worlds. In this paper, I shall adopt the hierarchical possible world semantics, i.e. the so-called 'hi-world semantics', developed in Tsai (2012) and try to provide a unified treatment of the logic of conditionals which, to a greater extent, catches the essence of everyday conditionals, indicative and subjunctive alike. Such semantics of conditionals not only is simpler than that of Stalnaker and Lewis but also sticks to Lowe's insight of using strict conditional as the backbone of a conditional. This in effect shows that Heylen and Horsten's negative result has not been conclusive. As a matter of fact, their analysis fails right at the beginning when they assume that

... it would scarcely be imaginable that the correct interpretation of conditionals essentially involves nested modalities. The resulting readings would be just too complicated for humans to use in ordinary reasoning. (Heylen and Horsten 2006: 540)

As we shall see soon, the hierarchical structure of a hi-world, consisting of different levels of case-classes, can play an essential role in our understanding of modality and conditionals.

¹ See, for instance, Stalnaker (1968) and Lewis (1973).

Two simple yet insightful ideas, due to O. Becker and F. P. Ramsey respectively, shall be the two pillars of our unified semantics for a language that contains conditionals. With Becker's insight, one can, through recognizing worlds of different levels, avoid the paradoxes of material implication, and with Ramsey's Test, we would not fall easy prey of the paradoxes of strict implication.

In the next section, we shall sketch the basics of Becker's semantics and the hi-world semantics, regarded as an alternative to the Kripkean semantics, and use it as the default semantics for our subsequent interpretation of modal operators. This by no means suggests that Kripke's semantics is in any way inferior to the Beckerian semantics. It is just that the account of conditionals that we will be proposing can be more straightforwardly discussed in Beckerian terms.

2. *Becker's Semantics and the Hi-world Semantics*

In Becker (1952), a "statistical interpretation of modal logic" was formulated in terms of cases and case-classes in such a way that a non-modal sentence P was to be evaluated against a case, while a primitive modal sentence (such as $\Box P$ and $\Diamond P$) and an iterated modal sentence (such as $\Box\Box P$ and $\Diamond\Box P$) were to be evaluated against a first level *case-class*² (i.e. a set of cases) and a second level case-class (i.e. a set of first level case-classes) respectively to yield a truth value. And the semantics is set up in such a way that $\Diamond\Box P$ is true with respect to a set U^2 of case-classes provided that among case-classes of U^2 , there is at least one case-class U^1 such that P is fulfilled in all cases contained in it. Higher degree situations can be worked out in the same spirit through induction: degree n modal sentences, i.e. iterated modal sentences with n modal operators, are to be evaluated against a level n case-class, where a level k case-class is a set of level $k-1$ case-classes, and a level 0 case-class is simply a case. A level 0 case can be seen, if one prefers, as a possible world, or, more properly, a plain world. A possible interpretation of the set U^1 is that it consists of all possible worlds consistent with one's present knowledge about the actual world. And, contrary to what some possible-world theorists would have said, this interpretation suggests that *counterfactual* possible worlds might not reside in U^1 , but rather reside in some subsets of U^2 . This in effect introduces a stratification into the realm of possibility.

A sentence of the form ' p or q ' can be concerned with two different kinds of entities. It may be saying something about a plain world w , claiming that the world is in the state prescribed by the sentence, or it may be saying something about a set U of possible worlds—claiming that each of those possible worlds is in the state prescribed in the earlier sense. So, the disjunction ' p or q ' can be translated either into $p\vee q$ or $p\vee q$, which abbreviates $\Box(p\vee q)$, and be evaluated against w and U

² Copeland's translation of *Fallklasse* is adopted here. See Copeland (2002).

respectively. The fact that no one would think ‘I live on Earth or I shall be assassinated by a Martian’ is true, while everyone would accept that ‘I live on Earth or I live on Mars’ is true does suggest that there is some subtle mechanism that drives us to take the U -reading for the former and the w -reading for the latter. So far as the semantics of a formal language is concerned, however, we do not need to know exactly how that mechanism works—we only need to acknowledge the existence of these two readings and know that they can be expressed in terms of connectives \vee and $\sqrt{}$ respectively.

Let us illustrate this phenomenon further with the *direct argument* discussed in Stalnaker (1975: 269).

(P) *Either the butler or the gardener did it.*

∴ (C) *If the butler didn't do it, the gardener did.*

Stalnaker elaborates on his pragmatic account and claims that the argument is indeed a *reasonable inference* but it is *invalid* nonetheless, so the validity of the following argument

(P1) *The butler did it.*

∴ (C) *If the butler didn't do it, the gardener did.*

would not follow from the apparent validity of ‘P1/∴P’. But, for us, it is only a matter of what reading— w -reading or U -reading—a speaker tends to have in mind for each of the sentences involved in the argument. The following are some of the possibilities, where $p \rightarrow q$ here abbreviates $\Box(p \supset q)$.

1. [P1- w ; P- w ; C- w]
B/∴.B \vee G valid, B \vee G /∴.¬B \supset G valid, and B /∴.¬B \supset G valid;
2. [P1- w ; P- w ; C- U]
B/∴.B \vee G valid, B \vee G /∴.¬B \rightarrow G invalid, and B /∴.¬B \rightarrow G invalid;
3. [P1- w ; P- U ; C- U]
B/∴.B \sqrt{G} invalid, B \sqrt{G} /∴.¬B \rightarrow G valid, and B /∴.¬B \rightarrow G invalid;
4. [P1- w ; P- w ; P- U ; C- U]
B/∴.B \vee G valid, B \sqrt{G} /∴.¬B \rightarrow G valid, and B /∴.¬B \rightarrow G invalid;

Clearly, only Case 4 captures the intuition of Stalnaker’s reader—before the notion of “reasonable inference” were made available—but it involves a subtle shift in the interpretation of the disjunction ‘either the butler or the gardener did it’ from one argument to another.

This is a promising result for Becker’s semantics, but Becker’s semantics actually faces a serious challenge that partly explains its poor reception in the early days of possible world semantics. This is the inconsistency in the process of evaluating sentences: sometimes you call for a world w , and sometimes you call for a set U of possible worlds. On the face of it, this separates the set of sentences into two subsets, w -sentences and U -sentences. But, the problem lies deeper—Becker’s

semantics cannot cope with sentences such as $\Box P \wedge P$, which apparently is neither talking merely about w nor merely about U . Fortunately, this problem can be solved with the introduction of hi-worlds. As the hi-world semantics will play an essential role in this paper, we shall sketch it here for easy reference and the reader is referred to Tsai (2012) for more details.

Let the language \mathcal{L} of propositional modal logic be defined by the following BNF:

$$\varphi := p_i \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid (\varphi \vee \varphi) \mid (\varphi \supset \varphi) \mid (\varphi \equiv \varphi) \mid \Box\varphi \mid \Diamond\varphi$$

where p_i is any atomic formula. A model M for \mathcal{L} consists of a non-empty domain set D , together with an interpretation function I which assigns a subset $I(p_i)$ of D to each atomic formula p_i . Intuitively, one can think of an element w of D as a Kripkean possible world, but to avoid confusion, we shall refer to it merely as a *plain-world*. Now, a *hi-world* s is of the form (U^0, U^1, U^2, \dots) , where U^0 is a plain-world w , and U^i is a level i world, i.e. an element of $(\mathcal{P})^i(D)$, where \mathcal{P} is the power set operator and $\mathcal{P}(A) = \mathcal{P}(A) \setminus \{\emptyset\}$. In short, a hi-world s is an element of $\prod_{i=0}^{\infty} (\mathcal{P}^*)^i(D)$. A hi-world t is a *sub-hi-world* of s provided that $\pi_i(t) \in \pi_{i+1}(s)$ for all $i \geq 0$, where π_i is the projection into the i -th component. The interpretation $\|\alpha\|_M$ of a formula α with respect to M is given by $\|\alpha\|_M = \prod_{i=1}^{\infty} U^i$, where $U^1 = I(\alpha)$ and $U^i = (\mathcal{P}^*)^i(D)$ for $i > 1$.

The hi-world semantics can then be given by

i) If α is a formula, then

$$\|\neg\alpha\|_M = \prod_{i=0}^{\infty} (\mathcal{P}^*)^i(D) \setminus \|\alpha\|_M$$

$$\|\Box\alpha\|_M = \{s \in \prod_{i=0}^{\infty} (\mathcal{P}^*)^i(D) \mid t \in \|\alpha\|_M \text{ for all sub-hi-worlds } t \text{ of } s\}$$

$$\|\Diamond\alpha\|_M = \{s \in \prod_{i=0}^{\infty} (\mathcal{P}^*)^i(D) \mid \text{there is a sub-hi-worlds } t \text{ of } s \text{ such that } t \in \|\alpha\|_M\}$$

ii) If α and β are formulas, then

$$\|\alpha \wedge \beta\|_M = \|\alpha\|_M \cap \|\beta\|_M$$

$$\|\alpha \vee \beta\|_M = \|\alpha\|_M \cup \|\beta\|_M$$

$$\|\alpha \supset \beta\|_M = \|\neg\alpha \vee \beta\|_M$$

$$\|\alpha \equiv \beta\|_M = \|\alpha \supset \beta\|_M \cap \|\beta \supset \alpha\|_M$$

Interestingly, we can introduce $\alpha \vee \beta \equiv \Box(\alpha \vee \beta)$ and $\alpha \rightarrow \beta \equiv \Box(\alpha \supset \beta)$ to force the usual U -readings of disjunctions and conditionals that we discussed earlier.

3. Ramsey's Test—Imposing the Antecedent

In a footnote to his paper 'General propositions and causality', Ramsey famously says the following about conditionals,

If two people are arguing 'If p will q ?' and are both in doubt as to p , they are adding p hypothetically to their stock of knowledge and arguing on that basis about q ; so that in a sense 'If p , q ' and 'If p , $\neg q$ ' are contradictories.

(Ramsey 1990: 155, footnote 1)

This passage is usually referred to as ‘Ramsey’s Test’ in the literature. However, to my knowledge, Ramsey did not call it a *test* himself, and it is indeed not merely a test. It can provide us with a general truth condition for conditionals, and captures some central features of conditionals that have been ignored by many theorists and hence caused many unnecessary conceptual difficulties concerning conditionals.

In this passage, Ramsey imagines that two people are disputing about the truth of a conditional “If p then q ” and then explains to us what these people actually do: they are adding p hypothetically to their stock of knowledge and arguing on that basis about q . In other words, Ramsey, in effect, outlines a truth condition for the conditional, and the truth condition roughly takes this form: a conditional ‘If p , q ’ is true for S provided that S adds p hypothetically to her stock of knowledge and on that basis accepts q .³ Therefore, if we stick to the framework of a truth-conditional semantics—that the meaning of a sentence is exhausted by its truth condition—then Ramsey’s Test, in short RT, amounts to the core of a theory of conditionals.

Now, if we are indeed concerned with the truth of a conditional of the form ‘If p , q ’, and are *unsure* about how the truth is to be determined—or we would not need RT in the first place—then we should take every care to ensure that in the process of carrying out RT, no other conditionals are employed. For otherwise RT would become a circular process that leads us nowhere—it invites a conditional to explain the conditional, while the meaning of the conditional introduced remains unexplained. Before spelling out what Ramsey really suggests, let us first look at a recent debate concerning RT so as to know how easily RT can be misinterpreted.

In Chalmers and Hájek (2007), the authors claim that ‘Ramseyan and Moorean principles entail that rational subjects should accept that they have the epistemic powers of a god’, in short, Ramsey + Moore = God. Barnett (2008) on the other hand claims that Chalmers and Hájek have interpreted Ramsey’s Test incorrectly, and that, when suitably interpreted, Ramsey + Moore \neq God. I shall show that both accounts involve circular explanation of conditionals, so that their arguments in support of their respective results can simply be discarded.

The positions of Chalmers and Hájek (2007) and Barnett (2008) can be summed up as follows. According to Chalmers and Hájek, Ramsey’s Test amounts to the following.

- (0) [C&H’s Ramsey] ‘if p then q ’ is acceptable to a subject S iff, were S to accept p and consider q , S would accept q .

C&H’s Ramsey together with Moore’s rationality principles would yield that, for a rational subject,

³ Apparently, such a truth condition suggests that people could disagree upon the truth of a conditional. However, this is not a drawback of the account. Rather, it reflects the true nature of real-life conditionals.

[Moore #1] If p , then I believe p ,

and

[Moore #2] If I believe p , then p ,

are acceptable and thus we get

Ramsey + Moore = God. — (※)

On the other hand, after introducing a notion of General Acceptability to account for the difference between acceptance and acceptability, Barnett arrives at the conclusion that Ramsey’s Test should rather be interpreted as

(0’) [Barnett’s Ramsey] ‘if p then q ’ is acceptable to a subject S iff, were S to hypothetically accept p and, on that basis, consider q , S would, on that basis, accept q .

According to Barnett, with this interpretation of the Ramsey Test, Moore #1 and Moore #2 are no longer acceptable, and we need not be bothered by the absurd result (※).⁴

Now, the true spirit of RT is to pin down the evaluation process of a conditional in terms of no other conditionals, yet while (0) introduces ‘were S to’ into its description of the process, (0’) complicates the matter even further by coming up with the phrase ‘were S to hypothetically’. Note that, generally, ‘were S to’ in itself starts a counterfactual conditional, which can be roughly paraphrased as ‘if $S...$ ’⁵ If RT is supposed to explain for us what ‘if ... then ...’ means, how can the very notion itself be employed to do the job? These authors have indeed gone along the opposite direction that Ramsey suggests us to go. They make RT entirely dispensable: if we can understand conditionals perfectly well, then what is the point of inviting RT into play in the first place? The absurdity of C&H program (Barnett’s is even more awkward) can be illustrated through the ironic equivalency of the following statements—it leads us to an infinite regression without explaining what ‘if ... then...’ actually means.

“‘if p then q ’ is acceptable to S_1 ’ is acceptable to S_2 ’ is acceptable to S_3 .
 ⇔ “‘if S_1 accepts p and considers q , then S_1 would accept q ’ is acceptable to S_2 ’ is acceptable to S_3 .
 ⇔ ‘if S_2 accepts ‘ S_1 accepts p and considers q ’ and considers ‘ S_1 accepts q ’, then S_2 would accepts ‘ S_1 accepts q ’ is acceptable to S_3 .
 ⇔ If S_3 accepts ‘ S_2 accepts ‘ S_1 accepts p and considers q ’ and considers ‘ S_1 accepts q ’ and considers ‘ S_2 accepts ‘ S_1 accepts q ’, then S_3 would accepts ‘ S_2 accepts ‘ S_1 accepts q ’.”

⁴ Barnett’s position is further stressed in Willer (2010: 292), where Willer tries to draw the reader’s attention to the fact that Ramsey “suggested that the antecedent is not accepted but only *hypothetically* added to what the agent believes to be true”.

⁵ Note that ‘were S to’ and ‘if S ’ behave differently so far as grammar is concerned. But we ignore this issue.

Alas, there is no way to get rid of the ‘if, then’. The reader, S_4 say, of the last of these sentences still has to figure out whether the conditional ‘**If...then...**’ is acceptable. Evidently, this is unlikely what Ramsey had in mind when he wrote down his famous footnote in question.

RT as a truth-condition for conditionals

One remarkable feature of RT is that Ramsey himself does *not* commit this fallacy of circularity. He uses the word ‘hypothetically’ so carefully that, on the one hand, one smells the flavor of a conditional through the employment of the term, and on the other hand, the evaluation process outlined in RT remains a declarative statement of the form ‘they are ... and ...’, the grasping of which does not presuppose the grasping of ‘if ... then ...’. Moreover, this allows us to have a truth-conditional semantics that can handle sentences with/without conditionals in a unified way.

To decide whether someone, S say, would assert ‘if p then q ’, Ramsey suggests that⁶

(R) S asserts ‘if p then q ’ iff S hypothetically adds p into her stock of knowledge and considers q and, on that basis, asserts q .⁷

Note that there is nothing conditional on the right hand of ‘iff’, and Ramsey has succeeded in providing us with a criterion for S ’s assertion of ‘if p then q ’. The nasty problem of ‘whether if p then q ?’ has now been turned into one concerning the mental reality of S , and the latter then provides us with a definite yes-no answer to the assertion of q given p .⁸ This is the key point of Ramsey’s proposal—shifting one’s focus from an entailment relationship between *world affairs* to an entailment relationship between *beliefs* of a person. Furthermore, we only need to know that there exists such a mental mechanism that would produce a yes-no answer to the conditional, not having to worry about what the detailed reasoning process of S actually is.

However, what do we mean by ‘*hypothetically* adding a belief p into one’s stock of knowledge’? Is it the same as ‘adding a belief p into one’s stock of knowledge’? Apparently not, because otherwise the term ‘hypothetically’ would be redundant. Nevertheless, the difference is subtler than we expect, and it will take me some time to explain it here.

Recall that in elementary logic, to prove the argument $r \therefore p \supset (p \wedge r)$, our friend S often use Conditional Proof as follows,

⁶ Note that on the left hand side of ‘iff’ we are using the word ‘accepts’, in contrast to the word ‘acceptable’ used on the left hand side of ‘iff’ in (0).

⁷ Some might object that in (R), I have used the term ‘iff’ which involves the notion of ‘if’ that I set out to explain, so I myself fall prey of the circularity problem. To this my reply is: i) ‘iff’ need not involve ‘if’ just as ‘=’ need not involve ‘≥’, ii) even if ‘iff’ involves ‘if’, so long as it is not used, as a meta-concept, in the definiens—the Right Hand Side of ‘iff’, that is—the definition is not guilty of circularity.

⁸ Some might object that asserting ‘if p then q ’ and asserting q are different things, so (R) cannot be right. However, it will be shown later that the one that asserts q is not, strictly speaking, the one that asserts ‘if p then q ’ in the first place.

- | | |
|-----------------------------|------------------------|
| 1. r | Premise |
| 2. p | Premise* |
| 3. $p \wedge r$ | 1, 2, Conjunction |
| 4. $p \supset (p \wedge r)$ | 2–3, Conditional Proof |

Now S has a premise r in her stock of premises to begin with, which makes her stock of premises consisting of only one premise. Then at step 2, she *hypothetically* introduces another premise p into her stock of premises. A key question to ask here is ‘how many premises does S have now?’ If the answer is ‘one’, then S is not entitled to use the second premise p in step 3. If the answer is ‘two’, then it contradicts the fact that S has only one premise. Furthermore, S has no right to introduce a new premise into her stock as she wishes. What happens?

The fact is that when S gets past step 2, she is posing herself as some other agent \hat{S} who has, in addition to all the beliefs that S has, in her stock of knowledge the belief p , and it is this \hat{S} who does the reasoning at steps 2 and 3, instead of S .⁹ And only when we get to step 4 does \hat{S} get the sack and S goes on alone to deal with things to come. In sum, *throughout* the proof, S has only one premise (and \hat{S} has two). It is just that at steps 2 and 3, we find the recruiting of \hat{S} helpful. One thing important to note here is that this individual \hat{S} has the belief p *intrinsically* rather than *hypothetically*. To be more explicit, at step 2 and 3, there are two individuals S and \hat{S} hanging around, and while the real S has the belief p hypothetically, the hypothetical \hat{S} has p intrinsically.

Now, back to (R), with the help of this individual \hat{S} , it is clear that the clause on the right hand side of ‘iff’ in (R) should be read as ‘ \hat{S} has p in her stock of knowledge and \hat{S} considers q and \hat{S} asserts q ’.¹⁰ Indeed we have a textual support from Ramsey (1990: p155) to hypothesize such an individual \hat{S} who has p in her stock of knowledge. The passage in question is this

So that in a sense ‘If p , q ’ and ‘If p , $-q$ ’ are contradictories.

Recall that S is by assumption in doubt as to p , so there is no reason why she would find $p \wedge q$ and $p \wedge -q$ contradictory—they may well be both false because p is a contradiction in itself. On the other hand, for an \hat{S} who has p in her stock of knowledge, $p \wedge q$ and $p \wedge -q$ are clearly contradictory. So the fact that Ramsey thinks ‘If p , q ’ and ‘If p , $-q$ ’ are contradictories suggests that he reasons as if there is such an \hat{S} who simply *has* p .

⁹ The introduction of \hat{S} turns the notion of hypothetical thinking into a concrete one. One can readily imagine that such \hat{S} ’s can be designed with the help of Artificial Intelligence.

¹⁰ The original qualification ‘on that basis’ in (R) serves to remind us that the assertion is made by the resulting \hat{S} rather than the S .

In the end, we can spell (R) out as follows

- (\mathfrak{R}) S asserts ‘if p then q ’ iff S poses herself as an \hat{S} that is like S in every aspect except that \hat{S} has p in her stock of knowledge, and \hat{S} considers q , and \hat{S} asserts q .

On the face of it, (\mathfrak{R}) is easy to understand and has great explanatory power, but the adding of p into one’s stock of knowledge alone will inevitably generate some nasty problem concerning one’s personal identity. We will defer the treatment of this complication until Section 5.

Alternatively, we can adopt the language of hi-world semantics and spell (R) out more explicitly as the following truth condition:

- (\mathfrak{R}^*) The conditional ‘if p then q ’ is true for a hi-world s iff there is a sub-hi-world s' of s such that p holds, and for any such s' , q holds as well.

The set \check{s}_p of all such sub-hi-worlds of s can be associated with the \hat{S} in (\mathfrak{R}) in the following sense. Let \check{s} denote the set of all sub-hi-worlds of s , then \check{s}_p is simply $\check{s} \cap \parallel p \parallel_M$.¹¹ For a non-modal p , this amounts to modifying the U^i of s into its intersection with the interpretation $I(p)$, that is $U^i \cap I(p)$, and obtaining a new hi-world \hat{s} . Thus s and \hat{s} correspond to S and \hat{S} of (\mathfrak{R}) perfectly. This correspondence, however, may not hold for an antecedent p that involves modality of different levels. When p is a sentence that mixes modality of different levels together, for instance, $p = (A \vee \Diamond B)$, then there may not exist an \hat{s} such that \check{s}_p corresponds to the set of all sub-hi-worlds of \hat{s} . In such cases, $\parallel p \parallel_M$ is not necessarily a product set, thus its intersection with the product set \check{s} , is not necessarily a product set. Therefore, the \hat{S} in (\mathfrak{R}) can only be thought of as a hypothetical individual who possesses the mindset \check{s}_p , which is concretely specified as a set of hi-worlds.

As a consequence, in terms of the language of propositional modal logic, Ramsey’s conditional ‘if p then q ’ can be translated as $\Diamond p \wedge \Box(p \supset q)$. I shall call this the *default* of a conditional. The $\Diamond p$ part plays a key role in the understanding of Ramsey’s conditional and it corresponds to the phrase ‘adding p hypothetically to their stock of knowledge’ in the Ramsey Test. In effect, $\Diamond p$ forces us to consider the possibility of p , while the $\Box(p \supset q)$ part requires us to restrict our attention to all those worlds such that p holds, and then we set our mind on q to see whether for all those worlds, q holds.

Recall that the paradoxes of strict implication take the following form. Were ‘if p then q ’ interpreted as the strict conditional $\Box(p \supset q)$, then

- S1 not $\Diamond p / \therefore$ if p then q
 S2 $\Box q / \therefore$ if p then q

¹¹ The reader is referred to Section 2 for the meaning of a sub-hi-world and the definition of $\parallel p \parallel_M$.

are both valid argument forms, yet the following typical instances of them are clearly invalid,

- Ex1 / \therefore *If 1=2 then I am happy.*
 Ex2 / \therefore *If I am happy then 1<2.*

Now let us see how our candidate $\Diamond p \wedge \Box(p \supset q)$ fares on these matters. One interesting fact to note first is that this candidate was actually bypassed by Lowe in Lowe (1995). It can avoid the first paradox of strict implication right away, but its alleged failure to cope with the second paradox of strict implication has led Lowe to turn to another candidate. It is not difficult to see that with this interpretation, S2 is indeed valid provided that p is possible. However, for S2 to be a non-trivial argument, the conclusion ‘if p then q ’ will have to be understood differently, namely in a subjunctive mode, and S2 then becomes invalid. As we need to resort to hi-worlds to have a better grasp of the subjunctive mode, we will postpone the detailed treatment to the next section.

However, for the moment, we can at least observe that Ex2 is not an instance of S2 at all. In other words, the utterer of Ex2 may not reckon $1 < 2$ as a necessary truth. So the absurdity of Ex2 may not entail the invalidity of S2. Hitchcock (1998) suggests that ‘ p entails q ’ can be interpreted as $\Box(p \supset q) \wedge (\Diamond p \vee \Diamond \sim q)$ to avoid both PSI1 and PSI2, but it does not work for the cases exemplified by Ex 1 and Ex 2. Another candidate $\Box(p \supset q) \wedge \Diamond p \wedge \Diamond \sim q$ would render both S1 and S2 invalid as desired, however, the adding of $\Diamond \sim q$ is too strong a requirement, because it would make the truism ‘if $1=1$ then $1=1$ ’ false.

In addition to the PSI, Lowe is also worried about the fact that the interpretation $\Diamond p \wedge \Box(p \supset q)$ would have deemed the following statement invalid, while it is surely a mathematical truism.

- (#) If n were the greatest natural number, then there would be a natural number greater than n .

This, in the end, leads Lowe to propose interpreting ‘if p then q ’ as $\Box(p \supset q) \wedge (\Diamond p \vee \Box q)$.

Indeed, (#) belongs to a category of conditionals that deserve more of our attention¹² but Lowe’s approach solves it at the price of accepting the second paradox of strict implication. The introduction of the condition $\Box q$ for cases where p is not possible does avoid the problem of (#), but it runs against the direction suggested by the second paradox of strict implication. On the other hand, the fact that (#) should be deemed valid can indeed be dealt with naturally and beautifully by the hi-world semantics as will be discussed in the next section. More specifically, when $\sim \Diamond p$ is true or $\Box q$ holds, we need not concede right away that ‘if p then q ’ is true, hence Ex1 and Ex2 can be invalid while (#) is valid. Ramsey’s idea and Tsai’s hierarchy together allows us to deal with these situations in a unified way which is consistent with our intuition.

¹² The reader is referred to Heylen & Horsten (2005) for a revised version of it.

4. Conditionals and hi-worlds

If $\diamond p \wedge \Box(p \supset q)$ in itself would not account for the validity of (#), how are we to cope with this situation?

The central idea of the hi-world semantics is that a hi-world s consists of worlds U^i of different levels. Now, the Default of a conditional $\diamond p \wedge \Box(p \supset q)$ is clearly to be evaluated against a level-1 world U^1 . However, if we have $\sim \diamond p$, then no plain world w in U^1 is such that p holds, thus seemingly an essential step in Ramsey's Test, namely that of 'adding p ', cannot be carried out. In this case, do we simply say that the conditional 'if p then q ' is false? Surely not! Re-examining Ramsey's Test more closely, we would find that Ramsey simply assumes that 'adding p ' is always a possible action. In other words, if we are to be true to Ramsey's spirit, then we need to be prepared to give truth values to a conditional 'if p then q ' even for cases where p is not possible.

So, we are challenged with a Mission Impossible, are we not? Certainly not. Recall that a hi-world s can be thought of as a string of worlds of all levels (U^0, U^1, U^2, \dots). Now, $\sim \diamond p$ says that there is no level-0 world w in the level-1 world U^1 such that p holds.¹³ However, that does not mean we have no way to conceive of a plain world w in the entire hierarchy of s such that p holds. As a matter of fact, beside the most natural place to look for such worlds, namely U^1 , the next candidate that comes to our mind is certainly U^2 . So, when we are given $\sim \diamond p$, and prompted by Ramsey's command to "add p to our stock of knowledge", what we need to do is simply imposing $\diamond \diamond p$, which claims that there is a level-1 world U^1 in the level-2 world U^2 such that there exists a plain world w in U^1 such that p holds.¹⁴

This insight leads us to the following unified account: a conditional 'if p then q ' in a natural language can be translated into one of the following sentences in the language of propositional modal logic,

Unified $(p \supset q)$ or $[\diamond p \wedge \Box(p \supset q)]$ or $\{\sim \diamond p \wedge [\diamond^2 p \wedge \Box^2(p \supset q)]\}$

where \diamond^2 and \Box^2 are the shorthand for $\diamond \diamond$ and $\Box \Box$ respectively. Note that, as we have seen in Section 2, the w -reading $p \supset q$ is seldom what we have in mind when we utter a conditional. Furthermore, recall that, in Tsai (2012), under the mild assumption that $U^i \in U^{i+1}$, for $i \geq 0$, we have that $\Box \alpha$ entails α , for any α , and for most people this is a quite natural assumption—if something happens in all possible worlds then it certainly would happen in this world as well. It is not difficult to see that if we adopt this assumption and disregard the primitive w -reading $p \supset q$, then **Unified** simply reduces to

¹³ Recall that U^i is determined by one's stock of knowledge in such a way that if you *know* that α , then your U^i can only consist of plain worlds such that α .

¹⁴ Gauker (2005) has introduced similar concepts, such as plain contexts and multi-contexts, into his account of conditionals. But unlike Gauker's account, our account here is more robust, not assuming any notion similar to Lewis' sequence of centered spheres.

Core $\{\diamond p \wedge \Box(p \supset q)\}$ or $\{\diamond^2 p \wedge \Box^2(p \supset q)\}$

One might wonder what happens to the truth of a conditional ‘if p then q ’ if its antecedent p is not only such that $\sim\diamond p$ holds but also $\sim\diamond^2 p$ holds. Should **Core** render the conditional false then? As a matter of fact this is not an option at all, because, as suggested by Ramsey, when we employ conditionals to convey our thoughts, the state of affair described by the antecedent has to be conceivable for us. If an antecedent is not only impossible but also necessarily impossible,¹⁵ then in practice it amounts to inconceivability for any utterer of the conditional, and we should regard it as *meaningless*—the predicate ‘true’ is simply inapplicable to such a conditional—rather than regard it as false. For instance, if you can conceive of the existence of a round square in your U^2 then the conditional ‘if there is a round square then geometry has to be rewritten’ is meaningful and can be either true or false; but if you cannot, then the conditional is simply meaningless for you. This issue can be further illustrated by how we answer a query raised by an anonymous reviewer for an earlier version of this paper. According to the reviewer, while ‘if $\Box A \wedge \sim A$ then $A \wedge \sim A$ ’ is seemingly true, the present account rules it as false because we have $\sim\diamond\diamond(\Box A \wedge \sim A)$. The point that I would stress here is that, given the mild assumption that $U^i \in U^{i+1}$, for $i \geq 0$, we indeed have $\sim\diamond(\Box A \wedge \sim A)$. But, to make sense of the conditional in question, we *have to* impose $\diamond\diamond(\Box A \wedge \sim A)$ anyway. And while it seems, at least for the reviewer, that $\sim\diamond\diamond(\Box A \wedge \sim A)$ holds naturally, it is not necessarily the case. While $\diamond p$ concerns possibility, $\diamond\diamond p$ concerns the *possibility of possibility*, and to impose $\diamond\diamond p$ we would naturally drop, if necessary, the possibility scheme at the lower, that is $\diamond p$, level, and this is how the present scheme works.¹⁶

Now, **Core** offers two possible reading for a conditional. In a real life discourse, a glimpse at a conditional ‘if p then q ’ usually suffices to make us opt for one of them,

Default $\diamond p \wedge \Box(p \supset q)$ when p is deemed possible,

Subjunctive $\diamond^2 p \wedge \Box^2(p \supset q)$ with p is deemed impossible.

Interested readers can check for themselves that the famous pair of Oswald-Kennedy conditionals can be explained in terms of these two readings. In the indicative mood, due to the *known* fact that the assassination has indeed happened, our U^1 does not contain any plain world in which Kennedy has not been assassinated; while in the subjunctive mood, our U^1 consists only of plain worlds in which Kennedy was assassinated by Oswald, so in order to impose the antecedent we are forced to resort to some plain worlds in some U'' of U^2 in which Oswald did

¹⁵ Theoretically one might consider pushing the possibility to still higher levels, for instance, taking into account $\diamond^3 p$. But in practice, reaching $\diamond^2 p$ suffices for our everyday purposes.

¹⁶ In this example, drop the condition that all sub-hi-worlds should satisfy the condition $U^i \in U^{i+1}$, for $i \geq 0$ as well.

not assassinate Kennedy and to see whether in those plain worlds Kennedy would still be assassinated at all—by others, of course.

We have seen previously that the Default reading renders (#) false while our intuition deems it true. With the Subjunctive reading at hand, however, we can easily see that (#) can now be true if we could imagine the existence of a *pseudo*-mathematical system in which there is a greatest natural number n , and for such pseudo-mathematical systems we can always find a natural number greater than n .¹⁷

With this, the truth of the following two conditionals would agree with a layman's intuition that (a) is false while (b) is true

- (a) *If $3^8=2187$ then $3^8=2187 \times 3$.*
 (b) *If $3^8=2187$ then $3^9=2187 \times 3$.*

According to proponents of strict implication, the mathematical truth that $3^8=2187 \times 3=6561$ makes the antecedent $3^8=2187$ necessarily false, so both (a) and (b) are true. But an evaluator of these conditionals may not be aware of $3^8=6561$, so her U^I may indeed contain worlds in which $3^8=2187$ holds, so that the conditionals are not deemed true automatically. Even if the evaluator happens to be an expert in numbers who knows that $3^8=2187$ is necessarily false, according to Ramsey's requirement of forcing the antecedent, he still has to force himself to accept $\diamond^2 p$, and the truth values of (a) and (b) are still to be decided pending on whether $\Box^2(p \supset q)$. As a result, we obtain the following interesting table.

	Strict implication $\Box(p \supset q)$	Lowe $\Box(p \supset q)$ $\wedge(\Diamond p \vee \Box q)$	Hitchcock $\Box(p \supset q)$ $\wedge(\Diamond p \vee \Diamond \sim q)$	Default $\Box(p \supset q)$ $\wedge \Diamond p$	Core	Intuition
a)	T	T	T	F	F	F
b)	T	F	F	F	T	T

Note that **Core** differs from the other candidates in that it allows us to resort to second order modalities. To further appreciate the power of **Core**, let us consider the following example of iterated strengthening of the antecedent

- 1) *If John wins the race then Jane will be happy.*
- 2) *If John wins the race and dies of a heart attack immediately afterward, then Jane will be happy.*

¹⁷ It may contain results that are inconsistent with present day mathematics. The lesson one learns from (#) may be this: given that such pseudo mathematical systems are all inconsistent—the antecedent is inconsistent with present day mathematic—we conclude that it is impossible for the antecedent of (#) to be true, hence the Subjunctive reading is justified.

It takes the following form

- 1') If A then D ,
- 2') If A and B then D .

According to the MI and SI accounts of conditionals, they should be translated as $A \supset D$, $(A \wedge B) \supset D$, and $\Box(A \supset D)$, $\Box((A \wedge B) \supset D)$, respectively. And for both accounts, the truth of 1') guarantees that of 2'). But in practice, it is very likely that one would assert 1) and deny 2). So we need new interpretations of 1') and 2'). According to the account outlined earlier, they can be translated as

- 1) $\Diamond A \wedge \Box(A \supset D)$ w.r.t. $s=(U^0, U^1, U^2, \dots)$
- 2) $\Diamond^2(A \wedge B) \wedge \Box^2((A \wedge B) \supset D)$ w.r.t. $s=(U^0, U^1, U^2, \dots)$

The fact that 2) is thought to be false comes in two steps. First, the Default reading of 2'), namely $\Diamond(A \wedge B) \wedge \Box((A \wedge B) \supset D)$, is inapplicable because in one's mind, no world in U^1 is such that $A \wedge B$ —otherwise he or she would not assert 1) in the first place—so the first conjunct is false right away. Second, the Subjunctive reading $\Diamond^2(A \wedge B) \wedge \Box^2((A \wedge B) \supset D)$ of 2') asks us to search in U^2 for some U^1 such that it contains a plain worlds w for which $A \wedge B$ holds, and then see whether all such worlds w 's in all such U^1 's are such that D . As it is usually not the case that Jane would be happy in those circumstances, we would regard the Subjunctive reading as false.

The merit of this account is that its prediction for the reverse Sobel sequence automatically conforms to our intuition. Evidently, when 'if A and B then D ' and 'if A then D ' is uttered in this order, no indicative-subjunctive shift will be triggered: $\Diamond(A \wedge B)$ itself entails $\Diamond A$, and if $\Box((A \wedge B) \supset D)$ is false then so is $\Box(A \supset D)$.

An interesting pragmatic feature to note here is that the hi-world associated with a speaker can actually be a dynamic entity, in the sense that it can evolve with discourse. For example, a new U^1 may evolve from a previous U^2 after a discourse involving the strengthening of the antecedent. I believe that this can serve as the ground for a general account of Belief Revision. However, it will have to be dealt with elsewhere.

The other two invalid argument forms that are discussed in Lewis' account of counterfactual, namely Contraposition and Transitivity, can be similarly analyzed.

The Lewis-style example of "Contraposition"

$$\frac{\text{If Peter drinks, he won't get drunk.}}{\therefore \text{Had Peter gotten drunk, he didn't drink.}}$$

can be translated into the following argument and it is invalid.

$$\frac{\Diamond A \wedge \Box(A \supset \sim(A \wedge D))}{\therefore \Diamond^2(A \wedge D) \wedge \Box^2((A \wedge D) \supset \sim A)}$$

And the Lewisian example of "Transitivity"

If J. Edgar Hoover had been a communist, he would have been a traitor,
 If he had been born a Russian, then he would have been a communist,

 \therefore *If he had been born a Russian, he would have been a traitor.*

can be translated into the following argument¹⁸ and it is invalid.

$$\begin{array}{c} \diamond C \wedge \Box(C \supset T) \\ \diamond^2 R \wedge \Box^2(R \supset C) \\ \hline \therefore \diamond^2 R \wedge \Box^2(R \supset T) \end{array}$$

Finally, the belated resolution of the second paradox of strict implication is achieved by spelling S2 out as

$$\text{S2 } \Box q / \therefore \diamond^2 p \wedge \Box^2(p \supset q),$$

which is invalid again as expected.

5. A Further Complication Concerning the Thomason conditionals

Let us now come back to the charge of Chalmers & Hájek (2007) that Ramsey + Moore = God. For brevity and relevancy to the present paper, I shall consider issues associated with Moore #1 only, and leave Moore #2 untouched here—interested readers can work out the latter without much difficulty. In other words, I shall only be concerned with whether Ramsey + Moore = Omniscience. According to (R), for S to accept ‘if p then I believe p ’ is for S to hypothetically add p into her stock of knowledge and considers ‘I believe p ’ and accepts ‘I believe p ’. In terms of (\mathfrak{R}), Moore #1 should be stated as

[Moore #1 (\mathfrak{R})] For every rational being S who considers ‘if p then I believe p ’, S poses herself as an \hat{S} that is like S in every aspect except that \hat{S} has p in her stock of knowledge, and \hat{S} considers ‘I believe p ’, and \hat{S} accepts ‘I believe p ’.¹⁹

On the face of it, this seems obviously true, but in that case, aren’t we on our way to the claim that Ramsey + Moore = Omniscience? I now draw the reader’s attention to a subtle point that distinguishes two readings of [Moore #1 (\mathfrak{R})]. As I have pointed out earlier in Section 3, it

¹⁸ Despite that the first premise takes the form of a subjunctive conditional, it reckons the possibility that future evidences reveal that Hoover was indeed a communist, so should be translated according to the Default interpretation.

¹⁹ Alternatively, in terms of (\mathfrak{R}^*), we can think of S as a hi-world s and \hat{S} as capable of leading us to a set \check{s}_p of hi-worlds such that for any s' in \check{s}_p , p holds. And our job is to decide whether ‘I believe p ’ holds for s' as well. Recall that for a non-modal p , my notion of S and \hat{S} actually correspond to hi-worlds s and \hat{s} respectively. But, as the reader may not be familiar with the abstract notion of hi-world semantics, in the main text, I will continue using the metaphorical S and \hat{S} to explain what is going on.

is important to note that in the consideration of q —i.e. ‘I believe p ’—it is \hat{S} who does the reasoning, while S is a hypothetical individual who plays only a temporary role for the duration of the evaluation of the conditional. In that case, what is the referent of the ‘I’ that appears in q ? There are two possibilities

- i) (Autistic- \hat{S} reading) The ‘I’ denotes \hat{S} . In this case, Moore #1 is true. The term ‘autistic’ is adopted here because, during the reasoning process, \hat{S} thinks of the referent ‘I’ as sharing her mental state concerning the belief p .²⁰
- ii) (Realistic- \hat{S} reading) The ‘I’ denotes S . In this case, Moore #1 is false. Here \hat{S} thinks of the referent ‘I’ as S , i.e. the person she really was.

In sum, according to the autistic reading, when considering ‘if p then q ’, the subject in q shares the knowledge state of the hypothetical individual \hat{S} , while in the realistic reading, the subject in q is unaffected by the knowledge state of \hat{S} . And so long as we bear in mind the distinction between these two readings, we can easily see that while Moore #1 may hold for the autistic Ramsey, it does not hold for the realistic Ramsey. Given the absurdity of ‘Ramsey + Moore = Omniscience’, we conclude that when taken as a statement of *rationality*, ‘if p then I believe p ’ suggests an autistic reading (which is obviously true); yet when taken as a statement of *omniscience*, ‘if p then I believe p ’ suggests a realistic reading (which is unlikely to be true). And (9R) itself does not tell us how \hat{S} is to conceive of the “I”.

To illustrate this point further, consider the following pair of sentences.

- A1 If there is a bomb in this room, I will leave the room in no time.
- R1 If there is a bomb in this room, I will be blown into pieces.

Both sentences sound acceptable. However, if Ramsey’s remark that ‘if p , q ’ and ‘if p , $\neg q$ ’ are contradictories is correct, then A1 and R1 cannot be both true. But, how come we feel that both of them are true? It is because that in A1, \hat{S} reads the ‘I’ as \hat{S} , while in R1, \hat{S} reads the ‘I’ as S . And while \hat{S} knows that there is a bomb in this room, S does not. It is important to note that in considering a conditional, S seems to gain an additional pair of eyes, namely that of the hypothetical \hat{S} , yet it is \hat{S} who does the reasoning and \hat{S} can decide whether she would like the self-reflexive ‘I’ to refer to \hat{S} or S . Similar examples abound.

²⁰ It is helpful to recall a famous setting concerning autism. Suppose an individual \hat{S} was watching, through a semi-transparent glass window, what another individual S was doing in his room. Now, before S went out of the room to fetch some water, he put his fountain pen into #1 drawer of the desk and closed the drawer. While S was absent, someone sneaked into the room and took out the pen, put it into the #2 drawer, closed the drawer and then left. Now, when we ask \hat{S} ‘Which drawer would S open in order to retrieve his fountain pen after he comes back?’ an autistic \hat{S} would allegedly say #2, while an \hat{S} with second person perspective would have said #1.

- A2 If there is a ton of gold buried under my house, I would dig it up and become a rich man.
- R2 If there is a ton of gold buried under my house, I wouldn't have known it and would thus remain poor.
- A3 If tonight's lottery winning numbers are 1~6, I will pick these numbers on my ticket and win the lottery.
- R3 If tonight's lottery winning numbers are 1~6, I will have no chance of winning it.

Here, in the autistic A2 and A3, the 'T' refers to \hat{S} , who has the antecedent in her stock of knowledge, while in the realistic R2 and R3, the 'T' refers to S , who, under realistic conditions, does not have the antecedent in her stock of knowledge.

Note that the Thomason Conditional as discussed in van Fraassen (1980) can be similarly analyzed.

- (3) If my business partner is cheating on me, I will never know it.

Apparently, the autistic reading of the 'T' as \hat{S} in (3) renders it obviously false, and the realistic reading of the 'T' as S in (3) renders it likely true. However, the fact that the knowledge state of 'T' is the sole concern of the consequent suggests that the 'T' should be understood as S rather than \hat{S} —given that the antecedent is by default in the stock of knowledge of \hat{S} , pragmatics would deem it inappropriate for someone to deny this very fact in the consequent—thus (3) is likely to be true even in the realm of RT.

Now consider two analogous examples from Lewis (1986) and Jackson (1987) respectively:²¹

- (4) If Reagan works for the KGB, I'll never believe it.
- (5) If Reagan is bald, no one outside his immediate family knows it.

Note that Ramsey's Test itself does not, as many might have supposed, render (4) false directly, because, once again, we have two possible readings of the 'T' in (4). For (4) to be informative, the truth of the consequent should not be too obvious. Yet the autistic reading of (4), namely seeing the 'T' as \hat{S} , will, together with the assumption that \hat{S} is rational, render the consequent evidently false. Therefore, the only possible reading left is the realistic one. Similarly, for (5) to be possibly true rather than trivially false, \hat{S} can only adopt the realistic reading, not counting \hat{S} as someone outside Reagan's immediate family who knows of Reagan's baldness.

In sum, in (3) ~ (5), the realistic reading renders them possibly true, while the autistic reading renders them false, and the Non-Triviality criterion of the consequent suggests that the realistic reading is preferred. Our analysis can be generalized even further to cases where

²¹ According to Willer (2010), footnote 2, what is characteristic of such conditionals, is that "the consequent asserts the agent's ignorance or disbelief of the fact described in the antecedent".

the consequent does not involve the first person. Consider the following conditional,

(6) If Reagan works for the KGB, Ramsey will know it.

When *S* is trying to determine the truth of (6), she would first turn herself into an \hat{S} , and then decide whether \hat{S} 's Ramsey shares with \hat{S} the knowledge of the antecedent of (6). If the answer is 'yes', then she is having the autistic reading of Ramsey, i.e. forcing her Ramsey to know out of the blue the thing she herself knows. The use of the term 'autistic' is justified by the fact that \hat{S} imposes the knowledge of Reagan's being a double agent—part of \hat{S} 's mental state—onto Ramsey. If the answer is 'no', then the 'Ramsey' is read realistically. And I believe most people would, taking into consideration the Non-Triviality criterion I mentioned earlier, read (6) the realistic way and render it likely false.

Variants of (6) can admit an autistic reading as well. Consider, for instance, the following conditional,

(7) If Reagan works for the KGB, Ramsey will report it.

If the 'Ramsey' in (7) is read realistically, then there is no guarantee that he knows that Reagan works for the KGB, and then how on the earth can we expect him to report it? Nonetheless, many people would be happy to accept (7), taking into account the fearless character of Ramsey. That is to say, they tend to read the Ramsey in (7) as already possessing, as \hat{S} does, the knowledge of Reagan's being a double agent. I believe such an autistic reading is the default reading when one faces (7). In contrast to the Non-Triviality criterion that I have employed in the reading of (6) so as to make sure that (6) is informative rather than being a truism, here a Relevancy criterion ensures that the 'Ramsey' in (7) has the relevant background information for him to decide upon whether to report it or not.²²

As is evident from the analysis of this Section, our account set up in the previous sections alone cannot successfully resolve Chalmers and Hájek's challenge. However, with the discovery of this additional subtle distinction between autistic and realistic readings, such problems are solved in the end.

²² Finally, it is interesting to observe that (7) can admit a realistic reading still, and it can be easily illustrated by the following dialogue.

Q: If Reagan works for the KGB, will Ramsey report it?

A: No way!

Q: How come?

A: Ramsey died before Reagan turned nineteen!

6. Conclusion

Becker prompts us to distinguish between talks about cases and talks about case-classes, and Ramsey reminds us how conditionals are used in daily language, in particular, he stresses the key step of imposing the antecedent in the process of evaluating a conditional. Hi-world semantics incorporates Becker's insight into a single framework so that all sentences are evaluated against a hi-world s , which consists in a string of worlds of all levels (U^0, U^1, U^2, \dots). Indicative conditional 'if p, q ' and subjunctive conditional 'were p, q ' can then, usually, be expressed as $\Diamond p \wedge \Box(p \supset q)$ and $\Diamond^2 p \wedge \Box^2(p \supset q)$ respectively.

One further complication concerning conditionals was illustrated by a recent debate about whether Ramsey + Moore = God. Our account sheds light on this matter by revealing that an autistic reading and a realistic reading of the indexical 'I' in 'If p then I believe p ' have been employed by Chalmers and others to affirm the thesis of rationality and to assert the omniscience thesis at the same time. While rationality thesis generally concerns about the reaction of an individual S after she is aware of the affair p , i.e. it concerns \hat{S} rather than S , for S to be omniscient and know things in the future, she is expected to know them beforehand, and this amounts to reading the 'I' in 'If p then I believe p ' as S rather than \hat{S} . This subtle shift from reading 'I' as \hat{S} to reading it as S , subsequently turning an acceptable rationality statement to an unacceptable omniscience statement, then accounts for the confusion involved in Chalmers and Hájek (2007) and Barnett (2008).

Acknowledgements

The author would like to express his gratitude to all the anonymous reviewers for earlier versions of this paper. Their valuable suggestions and comments have helped shape the paper into the present form. This work was supported in part by the Ministry of Science and Technology, Taiwan [NSC 101-2410-H-715-001- and MOST 102-2410-H-715-001 -MY3] and by Mackay Medical College [1021B07 and 1031B03]. Thanks are due to them as well.

References

- Barnett, D. 2008. "Ramsey + Moore \neq God." *Analysis* 68 (2): 168–174.
- Becker, O. 1952. *Untersuchungen über den Modalkalkül*. Meisenheim/Glan: Westkulturverlag Anton Hain,
- Chalmers, D. and Hájek, A. 2007. "Ramsey + Moore = God." *Analysis* 67: 170–72.
- Copeland, B. J. 2002. "The Genesis of possible worlds semantics." *Journal of Philosophical Logic* 31: 99–137.
- van Fraassen, B. 1980. "Review of Brian Ellis, *Rational Belief Systems*." *Canadian Journal of Philosophy* 10 (3): 497–511.
- Gauker, C. 2005. *Conditionals in Context*. Cambridge, Mass.: MIT Press.

- Heylen, J. and Horsten, L. 2006. "Strict conditional: a negative result." *The Philosophical Quarterly* 56: 536–549.
- Hitchcock, D. 1998. "Does the traditional treatment of enthymemes rest on a mistake?" *Argumentation* 12: 15–37.
- Jackson, F. 1987. *Conditionals*. Oxford: Basil Blackwell.
- Lewis, D. 1973. *Counterfactuals*. Oxford: Blackwell.
- Lewis, D. K. 1986. "Postscripts to probability of conditionals and conditional probability." In *Philosophical Papers*, volume 2. Oxford University Press, New York: 152–156.
- Lowe, E. J. 1983. "A simplification of the logic of conditionals." *Notre Dame Journal of Formal Logic* 24 (3): 357–366.
- Lowe, E. J. 1995. "The truth about counterfactuals." *The Philosophical Quarterly* 45: 41–59.
- Ramsey, F. P. 1990. "General propositions and causality." In *Philosophical Papers*, edited by D. H. Mellor. Cambridge: Cambridge University Press: 145–163.
- Stalnaker, R. 1968. "A Theory of conditionals." In N. Rescher, *Studies in Logical Theory*. Oxford: Blackwell.
- Stalnaker, R. 1975. "Indicative conditionals." *Philosophia* 5: 269–286.
- Tsai, C-C. 2012. "The genesis of hi-worlds—towards a principle-based possible world semantics." *Erkenntnis* 76 (1): 101–114.
- Willer, M. 2010. "New surprises for the Ramsey Test." *Synthese* 176 (2): 291–309.

Plural Voting and J. S. Mill's Account of Democratic Legitimacy

IVAN CEROVAC

Department of Philosophy, University of Trieste, Trieste, Italy

This paper clarifies some of the contested ideas put forward by John Stuart Mill by analyzing the reasons and arguments Mill used to support them and demonstrating how these ideas and arguments supporting them are connected into a coherent system. Mill's theory is placed in wider explanatory framework of democratic legitimacy developed by Thomas Christiano, and is portrayed as a typical example of democratic instrumentalism—a monistic position that focuses on the outcomes and results of a decision-making process. Following this move, the focus is shifted on the understanding of political equality in Mill's political thought. I claim that, contrary to some contemporary interpretations, Mill's theory is based on a few fundamentally inegalitarian ideas. Finally, Mill's view on the role of experts in democratic decision-making is analyzed and compared with contemporary theories advocating democratic expertism—Mill's view is again portrayed as inegalitarian, both to the extent of setting political aims and creating methods for achieving these aims.

Keywords: Political legitimacy, plural voting, expertism, Mill, Christiano. epistemic democracy, division of epistemic labor.

1. Introduction

Many books and papers have been published criticizing Mill's plural voting proposal and analyzing its strengths and weaknesses, as well as trying to implement it (or criticize it) from the standpoint of contemporary western democracies (Baccarini 1993, Baccarini & Ivanković 2015, Brilhante & Rocha 2013, Gaus 2003, Miller 2003, Thompson 1976, Urbinati 2002). Unfortunately, it seems that Mill's original work is somehow neglected in favor of some notable interpretations, and the emphasis is sometimes placed on implementation of Mill's ideas in contemporary society without first analyzing and understanding the

justificatory process Mill carefully developed to support those ideas. This paper aims to clarify some of the contested ideas by analyzing the reasons and arguments Mill used to support them, as well as to emphasize how these ideas and arguments are connected into a coherent system. Furthermore, this paper tries to determine what is the role of plural voting in Mill's argument and how exactly does the plural voting proposal improve the epistemic quality of a democratic decision-making process.

First part of this paper sets Mill's account in the wider explanatory framework of democratic legitimacy developed by Thomas Christiano. Mill's view is portrayed as a typical example of democratic instrumentalism—a monistic position that focuses on the results of a decision-making process when discussing the legitimacy of the decisions produced by this process. Mill's understanding of political equality is discussed in the second part: by introducing Berlin's distinction between positive and negative liberties, I claim that Mill argued only for the equality of negative liberties. Positive liberties, those inherent to a participatory democratic process, are not to be equally distributed. Values of deliberative democracy and diverse perspectives are discussed in the third part. By building on these ideas, I point out why Mill believed that everyone should have a say in a decision-making process, though not everyone should have an equal say. Plural voting proposal satisfies perfectly the requirement Mill had in mind (unequal political power but participation of all in decision-making process) and is discussed in the fourth part of the paper. There I stress again Mill's allegiance to democratic instrumentalism by comparing his view on experts with the views of Thomas Christiano and Philip Kitcher. While Christiano and Kitcher advocate for equality in the process of setting up political aims (and give greater power to the experts only when discussing the implementation of the already set aims), I claim that Mill rejects the idea of equality both in the process of setting up aims and in the process of their implementation (though he has a different standard for identifying experts in these two domains). Some concluding remarks are presented in the final part of the paper, emphasizing the important role of public justification for Mill's view.

2. *Background*

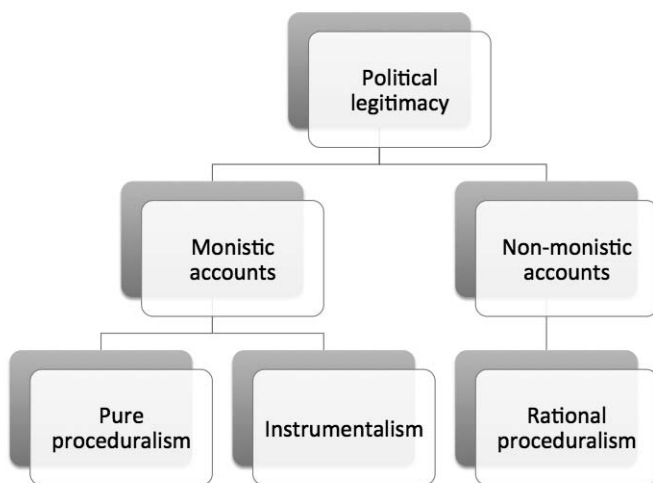
Whenever we try to justify or argue for certain form of government, we start by enlisting its virtues (Swift 2006). Contemporary political philosophy divides these virtues depending on whether they are the virtues of a decision-making process or of the final outcome produced by this process (Christiano 2004). In order to give an account of democratic legitimacy presented by John Stuart Mill, as well as to compare his position with those of contemporary defenders of epistemic democracy, one first has to clarify the criteria listed above and set Mill's view according to them.

Pure proceduralism focuses only on purely procedural qualities of a decision-making procedure when determining its legitimacy-generating potential. These purely procedural (sometimes called intrinsic) qualities are defined regardless of procedure's ability to produce certain goal or outcome—a decision-making procedure has legitimacy-generating potential because it embodies some important moral (or epistemic) qualities. Procedural fairness (i.e. giving every citizen an equal chance to participate in the decision-making process) can be one such purely procedural quality. A collective decision is thus legitimate if (and only if) it was produced by a fair decision-making procedure. Positions developed by Hannah Arendt (1967), Thomas Christiano (2008), Gerald Gaus (1996), Fabienne Peter (2011), Iris Marion Young (2000) and Robert Dahl (1989) are some examples of pure proceduralism.

Instrumentalism, on the other hand, focuses only on the instrumental qualities of a decision-making procedure when determining its legitimacy-generating potential. These instrumental qualities are defined by procedure's ability to reach a desired aim or outcome—a decision-making procedure has legitimacy-generating potential because of its ability to generate decisions with some substantial, procedure-independent quality. The ability to produce correct, true or just decisions can be one such instrumental quality. A collective decision is thus legitimate if (and only if) it was produced by a decision-making procedure that has tendency to produce correct or true decisions. Positions developed by Steven Wall (2007) and Richard Arneson (2003b), but also by Robert Talisse (2009) and Cheryl Misak (2000) are some examples of political instrumentalism.

We can try to justify democratic legitimacy by referring to one of these virtues, in which case we will be endorsing some monistic position, or we can try to justify democratic legitimacy by referring to both virtues, in which case we will be endorsing some non-monistic position.¹ The standard account of epistemic democracy put forward by David Estlund (2008) represents one such non-monistic position, focusing on both the fairness of the procedure and the qualities of the outcome.

¹ The distinction between monistic and non-monistic positions was first introduced by Christiano (2004).



3. *Mill's criteria for legitimacy*

Mill asserts that the best form of government is the one that best achieves the following two goals: (i) improving the virtue and intelligence of the people under its jurisdiction, and (ii) organizing the existing virtues and good qualities of the people in a way that promotes the long-run common good.

One criterion of the goodness of a government [is] the degree in which it tends to increase the sum of good qualities in the governed, collectively and individually; since, besides that their well-being is the sole object of government, their good qualities supply the moving force which works the machinery. The other constituent element of the merit of a government [is] the quality of machinery itself; that is, the degree in which it is adapted to take advantage of the amount of good qualities which may at any time exist, and make them instrumental to the right purposes. (Mill 1977a: 390–391)

The same two criteria reappear, more or less reformulated, throughout his entire work.

[Merit which any set of political institutions can possess] consists partly of a degree in which they promote the general mental advancement of the community, including [...] advancement in intellect, in virtue, and in practical activity and efficiency; and partly of the degree of perfection with which they organize the moral, intellectual and active worth already existing, so as to operate with the greatest effect of public affairs. (Mill 1977a: 392)

However, it seems that the basic and unifying criterion behind these two are beneficial consequences (Sandel 2009, Peter 2014). Therefore, though Mill enlists two criteria of good governance, his position remains monistic since both criteria regard only the consequences of a procedure (i.e. whether the procedure produces good decisions and

whether the procedure improves the qualities of citizens). Following Mill's utilitarian account characterized by the differentiation between higher-quality and lower-quality pleasures, the best form of government is to be understood as the one that produces maximal aggregate long-run utility (excellence-weighted pleasure).

The ideally best form of government is [...] the one which [...] is attended with the greatest amount of beneficial consequences, immediate and prospective (Mill 1977a: 404)

It is rather clear that Mill uses a procedure-independent criterion for evaluating the quality of the outcomes. A political decision can be good or bad regardless of the procedure that has produced it. This is particularly clear when Mill uses an epistemic argument to argue against despotic monarchy; even if there would be a wise benevolent despot, he would be unable to detect and promote the common good, as well as particular interests of different individuals, as efficiently as representative (democratic) government. A political decision is good or bad regardless of the procedure that has produced it; its quality is evaluated in the light of its consequences.

Mill adopts the instrumentalist position: a form of government is only legitimate if it produces the greatest possible amount of beneficial consequences. He avoids the common objections against utilitarianism by introducing the differentiation between higher-quality and lower-quality pleasures, as well as by strongly arguing that only by preservation of individual liberties we can maximize utility in a long-run.² However, his argumentation has an instrumental form; in order to be legitimate, a form of government has to improve intellectual and moral qualities of its citizens, as well as to organize them in such a way as to produce the best possible outcomes (Peter 2014).

What form of government will prove itself as the best depends on the people it is exercised upon. Tyranny will be the best form of government for barbarian tribes, since it will best improve their intellectual and moral qualities (e.g. teach them to obey the laws), as well as organize them in a manner they, because of the lack of discipline, would otherwise be unable to do themselves. Democracy is preferred to tyranny, but only when discussing developed societies where certain preconditions have already been met. This emphasizes the instrumental approach used by Mill: what form of government is legitimate depends on the type of society we want to apply it upon. Different forms of gov-

² Mill's famous essay '*On Liberty*' can be viewed as a unified attempt to argue in favor of individual liberty from the consequentialist (utilitarian) standpoint. All four reasons that explain why we should uphold individual liberty have an instrumental form—we should not silence the dissents because such an action would produce ill consequences for our society: we might be deprived of true or partially true belief, our own belief might harden into dogma and prejudice, and forcing the members of a society to embrace custom and convention is likely to deprive them of the energy and vitality for social improvement. For detailed argumentation see Mill 1879 and Sandel 2009.

ernment will yield different results when applied to different societies. Democracy is thus instrumentally justified: if we want to promote intellectual and moral qualities of individuals in our society, and if we want to organize them to produce the best possible outcomes, we should embrace democracy as a proper form of collective decision-making.

4. *Expertism and equality*

Mill's democratic instrumentalism can sometimes be mistaken for a weak kind of (epistemic) proceduralism: after all, Mill does not think that political decision is legitimate if and only if it has beneficial consequences. According to such view, whenever one has a reason to doubt the quality of consequences of a political decision, one could say that he does not recognize that particular decision as legitimate. This surely is not the result Mill had in mind. Furthermore, we could question the extent to which such view improves intellectual and moral qualities of the people involved. The decision-making procedure is very important for Mill—it has to be organized in such a way as to satisfy two criteria of good government, i.e. to improve the intellectual and moral qualities of people and to organize their potentials to maximize the quality of results. A decision is thus legitimate if it is a product of a good decision-making procedure. Though this might seem as a form of democratic proceduralism, we must note that the justification of the procedure is purely instrumental (Peter 2014). Mill does not find democracy superior to despotic monarchy because the former respects the equality of all the people involved, and the latter does not. His arguments for democracy have instrumental form; we should prefer democracy because it produces better outcomes, i.e. it is better in improving our moral and intellectual qualities, as well as in producing better decisions. Unlike Estlund and other philosophers who adopt non-monistic positions, putting emphasis on both the fairness of a procedure and the quality of results it produces, Mill's view is monistic—only the results are important.

Some might argue otherwise by stressing the importance of equality in Mill's political thought, especially in his famous essay *On Liberty* (Justman 1990). Though equality is indeed a very important idea for Mill, we must notice that in *On Liberty* Mill refers primarily on the idea of negative liberty, i.e. the area within which the subject—a person or group of persons—is or should be left to do or be what he is able to do or be, without interference by other persons (Berlin 1969). Mill's thoughts on positive liberty, i.e. his answer to the question what, or who, is the source of control or interference that can determine someone to do, or be, this rather than that (Berlin 1969), are quite different. Mill explicitly distinguishes the power that one has over oneself alone and the power one has over others:³

³ I thank David Miller for pointing this idea and encouraging me to analyse Mill's *Thoughts on Parliamentary Reform* in detail.

They say that everyone has an equal interest in being well governed, and that every one, therefore, has an equal claim to control over his own government. I might agree to this, if control over his own government were really a thing in question; but what I am asked to assent is, that every individual has an equal claim to control over the government of the other people. The power that suffrage gives is not over himself alone (*i.e. negative liberty*) it is power over others also (*i.e. positive liberty*): whatever control the voter is able to exercise over his own concerns, he exercises the same degree of it over those of every one else. Now, it can in no sort be admitted that all persons have an equal claim to power over others (Mill 1977b: 323)

It seems that equality does not play an important role in Mill's thoughts on *collective decision-making procedures*, though it still plays an important role with regard to the development of individual capacities (Baccarini 2013, Macpherson 2012); it is very important to ensure the equal protection of everyone's basic negative liberties (e.g. freedom of thought, speech, press and assembly), but equality should be rejected and opposed when discussing positive liberties. It should instead be replaced with competence and (non-equal) participation, because these are the key virtues needed to achieve better quality of political decisions.⁴

This particular idea is nicely implemented in the *plural voting* practice suggested by Mill. He indicates two motives for this proposal: (i) to prevent one group of people from being able to control the political process without having to give reasons in order to have sufficient support, and (ii) to avoid giving each person an equal chance to influence political decisions without regard to their merit, intelligence etc.

Yet in this stage of things, the great majority of voters [...] are manual laborers; and a twofold danger, that of too low a standard of political intelligence, and that of class legislation, would still exist in a very perilous degree. (Mill 1977a: 473)

⁴ To additionally stress this point, it might be useful to point out important differences between Mill's approach and the approach of those who base democratic legitimacy on the idea of equality (e.g. Thomas Christiano). Christiano builds his theory on a basic claim that human beings are authorities in the realm of value because (i) they are capable of recognizing, appreciating and producing value, and because (ii) their exercise of this authority is itself intrinsically valuable. Christiano further claims that equal status of persons is based on the fact that human beings all have essential the same basic capacities to be authorities in the realm of value (Christiano 2008). Mill, on the other hand, believes that people are obviously differently capable of appreciating intrinsic values (his version of 'higher pleasures' utilitarianism), and that differences in capacity should produce differences in status. This does not imply that those who are better educated should direct the private lives of those who are not (Mill clearly stresses this point in 'On Liberty'), nor should they have absolute power in political arena (this is pointed out in 'Considerations on Representative Government'). The underlying reason for this is not equality, however, but the idea that intellectual and moral qualities of all human beings should be cherished and improved, and that would be impossible if other people would direct our every action. This does not imply, however, the idea that everyone should have an equal say in a collective decision-making process.

Some scholars (often following the republican tradition) seem to believe that the main motive Mill had to suggest plural voting was to stop the tyranny of majority in a form of class legislation (Brilhante & Rocha 2013, Honohan 2002, Justman 1990, Miller 2000). After all, introducing plural voting and giving the educated (i.e. the minority of voters) more than one vote might look like an attempt to defend the republican value of non-domination (Pettit 1999). For example, Brilhante and Rocha claim “Mill would not have favored inequalities that implied undue power over others because this would undermine the autonomy that was a central value in his political philosophy. He advocated the plural voting system on the assumption that it would increase general happiness by preventing the tyranny of the majority” (2013: 62). The danger of too low standard of political intelligence is often neglected, and the entire plural voting proposal is regarded as a temporary solution Mill used ‘in [his] stage of things’, i.e. to answer the problem of British electorate in 19th century. However, there are good reasons to consider Mill’s plural voting account as a permanent solution. In fact, Mill’s own words oppose those who think that plural voting is only a temporary solution that should not be considered as an important part of his political thought.

I do not propose the plurality as a thing in itself undesirable, which, like the exclusion of a part of the community from the suffrage, may be temporarily tolerated while necessary to prevent greater evils. (Mill 1977a: 478)

It is clear, in fact, that Mill’s main reason for plural voting is not class legislation, the ‘greater evil’ from the previous quote. Even in a society where there is no fear of one class or group of people being able to control the political process without having to give reasons in order to have sufficient support, Mill would still opt for plural voting and against the equality of votes.

I do not look upon equal voting as among the things which are good in themselves, provided they can be guarded against inconveniences. I look upon it as only relatively good [...], but in principle wrong, because of recognizing a wrong standard, and exercising a bad influence on the voter’s mind. It is not useful, but hurtful, that the constitution of a country should declare ignorance to be entitled to as much political power as knowledge. (Mill 1977a: 478)

Mill was strongly influenced by the classical political philosophy, and his plural voting proposal can be seen as a combination of Plato’s epistocracy and Aristotle’s democracy. Following Plato, Mill emphasized the value of greater wisdom of the few, while following Aristotle he embraced the value of diverse perspectives for political decision-making. (Estlund 2003: 57) Though Mill never embraced Plato’s epistocracy (because it denied the value of diverse perspectives for decision-making, as well as because it was not compatible with the account of moral and intellectual improvement of the people), he considered the idea that competence should have greater weight than incompetence very appealing.

[...] that governing is not a thing which can be done at odd times, or by the way, in conjunction with a hundred other pursuits, nor to which a person can be competent without a large and liberal general education, followed by special and professional study, laborious and of long duration, directed to acquiring, not mere practical dexterity, but a scientific mastery of the subject. This is the strong side of the Platonic theory. (Mill 1978: 436)

When two persons who have a joint interest in any business, differ in opinion, does justice require that both opinions should be held of exactly equal value? If [...] one is superior to other in knowledge and intelligence, the judgment of a higher moral or intellectual being is worth more than that of an inferior: and if the institutions of the country virtually assert that they are of the same value, they assert a thing which is not. One of the two, as a wiser or better man, has a claim to a superior weight [...] (Mill 1977a: 473)

It seems clear that Mill argued for deliberative democracy on instrumental grounds; his plural voting proposal is an example of such argumentative strategy.

5. *Deliberative democracy*

One has to notice, however, that the reason why plural voting is introduced is not *only* to improve the quality of decisions produced by collective decision-making process. Mill emphasizes the educational role of democracy, and of the experts as well. Their influence will improve the quality of decisions, but it will also help common people further develop their intellectual and moral skills.

There are very good reasons not to believe that Mill adopted a form of elitism that could lead to epistocracy. We have indicated earlier that Mill recognizes the value of diverse perspectives, as well as the danger of class legislation. If we give overly exaggerated political power to certain group of people (even if they are experts), the danger of class legislation is reintroduced, and the value of diverse perspectives is lost. This value of diverse perspectives is best introduced through deliberation; though one can argue that even a form of aggregative democracy could take advantage of diverse perspectives and produce high-quality outcomes (e.g. Marquis de Condorcet, Kenneth Arrow), this is only one of the two goals of the good government. The other one, development of our intellectual and moral qualities, can only be achieved through deliberation.

Those who are supreme over everything, whether they be One, Few or Many, have no longer need of the arms of reason; they can make their mere will prevail; and those who cannot be resisted are usually to well satisfied with their own opinions to be willing to change them, or listen without impatience to anyone who tells them that they are in the wrong. [...] the one which develops the best and highest qualities is the position of those who are *strong enough to make reason prevail, but not strong enough to prevail against reason*. (Mill 1977a: 478–479)

Following this argumentation, one could be led to believe that the only reason for plural voting is to attain the balance between groups or

classes that would force them to deliberate instead of simply asserting their will, and the only reason for adopting deliberative procedures is to improve the moral and intellectual qualities of people engaged in deliberation. There are good reasons not to embrace this interpretation: though Mill's argumentation was aimed to maximize the individual liberty, this liberty can be limited when our actions have impact on lives of other individuals. As long as we make decisions that are within our private sphere, neither majority of the people nor (moral) experts should have an authority to limit our liberty. Things change, however, when our decisions influence other people beside us, just like all political decisions do. Giving greater power to the voice of an expert in such situation can be legitimate.

There would be no pretence for applying this doctrine to any case which could with reason be considered as one of individual and private right. In an affair that concerns only one of two persons, that one is entitled to follow his own opinion, however much wiser the other might be than himself. But we are speaking of things that equally concern them both; where, if the more ignorant does not yield to the guidance of the wiser man, the wiser man must resign to more ignorant. [...] No one but a fool, and a fool of peculiar description, feels offended by the acknowledgement that there are others whose opinion, and even whose wish, is entitled to a greater amount of consideration than his. (Mill 1977a: 473–474)

6. *The role of plural voting*

Mill is well aware of the defects any form of government might have. He points out that the worst defects a democratic government might face are its inability to produce good decisions and its tendency to be influenced by particular interests of dominant groups (Mill 1977: 436). Plural voting was introduced as a means to counter these defects: its main purpose was to ensure that the representative government produces high quality outcomes, and that no group has exclusive right to the benefits of social cooperation by the power of votes alone (and without having to deliberate and convince others to support the decision in question).

It is unclear, however, how exactly was plural voting proposal supposed to counter the first defect of democratic government, i.e. to ensure that the procedure produces good decisions. How was plural voting supposed to achieve its purpose? In their recent paper, Baccharini and Ivanković (2015) claim that plural voting proposal seriously threatens the quality of outcomes. It is unclear at which stage of the decision-making process does the epistemic value of plural voting help us create better policies and decisions. They analyze *the problem stage* (where political values are expressed and some problems are detected), *the proposal stage* (where the educated commission drafts laws and policies), and *the approval stage* (where the Parliament chooses to pass or reject a certain law proposed by the commission), and claim

that plural voting proposal does not bring epistemic value in any of the stages mentioned above. Similar objections are raised by Gaus (2008) and Peter (2012), who claim that it is very difficult to determine who the experts regarding some political issue are, and add that the relevant competences for making political decisions are often so widely dispersed that the (epistemic) distinction between citizens and experts is small and irrelevant, just like the (epistemic) distinction between procedures characterized by equal suffrage and those characterized by plural voting.

I do not want to argue that Mill's plural voting proposal has an epistemic value—all I want is to show why did Mill think it had epistemic value, and in which stage of the decision-making process did this epistemic value manifest itself. In order to answer these questions, we must first analyze the sophisticated structure of democratic government and the key stages of democratic decision-making process, as well as different concepts of expertise.

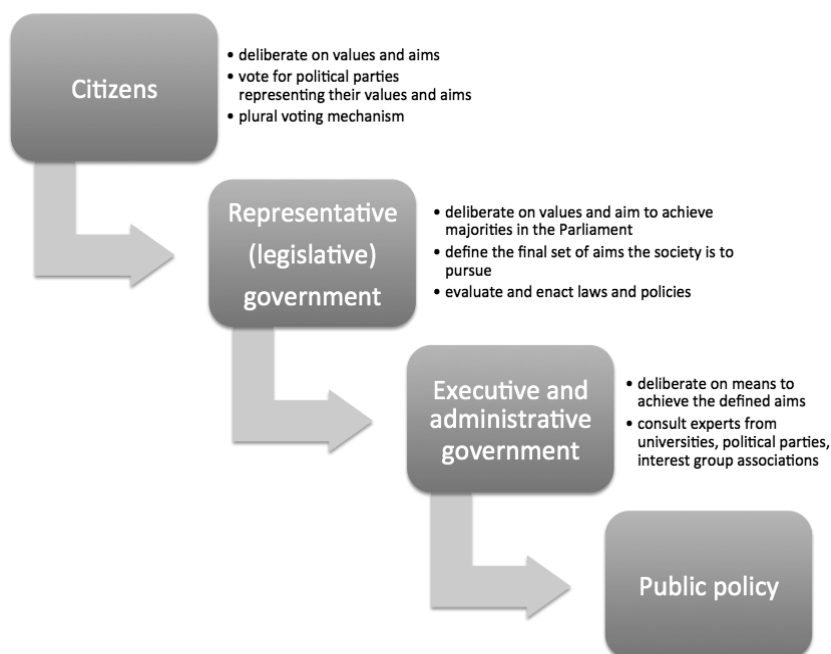
Thomas Christiano (2008) introduces a useful differentiation between technical and moral knowledge. *Technical knowledge* regards crafts, skills and disciplines like engineering, medicine, carpentry, physics, law or computer sciences. Most people can see this knowledge as useful and some educational institutions can be publicly seen as reliable sources of this knowledge. However, there is another kind of knowledge, one that regards what is right and what is wrong. This *moral knowledge* is about values and it is not as public as technical knowledge is, since we have a widespread disagreement on both the moral issues and the experts in morality (Christiano 2008). Mill agrees that the technical knowledge is probably more public than the moral knowledge, but unlike Christiano he thinks that we can still determine those whose '*opinions and even wishes*' should be given greater consideration. Mill does not set strict constraints on education (he does not insist that only philosophers, or only experts in political science or economics, have greater political power), nor does he name the exact profession one has to have in order to have a plural vote. His main idea is that people who have dedicated some time and effort to improving their intellectual and moral capacities are generally more capable of knowing what is more valuable in life (they are better acquainted with higher pleasures), and therefore are more capable of setting valuable aims for the society in general.

Mill firmly believed in the idea of epistemic division of labor and consequently, that laws and political decisions should be made by the most competent members of a society (i.e. experts). He saw division of labor as one of the central reasons for rejecting direct democracy, but nonetheless did not believe that parliament should make laws, public policies and political decisions. This task was appointed to small expert bodies (commissions), while it was the task of the Parliament to discuss and deliberate on proposed laws and decisions, as well as to accept or

refuse proposals made by such commissions (Mill 1977: 424). Unlike expert bodies, Mill did not think that the Parliament should be composed primarily of experts:

[Members of parliament] are not a selection of the greatest political minds in the country, from whose opinions little could with certainty be inferred concerning those of the nation, but are, when properly constituted, a fair sample of every grade of intellect among the people which is at all entitled to a voice in public affairs. Their part is to indicate wants, to be an organ for popular demands, a place of adverse discussion for all opinions relating to public matters, both great and small [...] (Mill 1977a: 433)

Therefore, considering the division of labor and a purely deliberative function of the parliament, Mill did not have in mind that plural voting will directly ensure more competent law-makers and policy-makers. The competences of law-makers and policy-makers can be similar both under monarchical and democratic rule (Mill 1977a: 438–439). Plural voting is introduced to give additional strength to opinions and even *wishes* of those better educated, and to increase the number of people representing these opinions and wishes in the parliament. If small expert bodies (commissions) are those who devise practical means (laws, policies, decisions) to achieve a desired political end, it is parliament who sets these political ends, and in setting them, the parliament represents the general public, but plural voting enables it to put a greater emphasis on those ends that well-educated people consider valuable (because their opinions are better represented in the parliament). Plural voting thus improves the quality of political decisions not by improving the technical process of finding best practical solutions to designated problems, but by improving the quality of political aims we as a society want to achieve. In other words, the epistemic value of plural voting is introduced primarily in the problem stage of democratic decision-making process. What shall we define as a problem in a society depends on the values and aims we want to pursue. For example, if we want to protect the traditional family with father as breadwinner and mother as caretaker (Kristol 1995), having a 40% unemployment rate will not be a serious political problem, as long as those unemployed are women. Similarly, if our political aim is full employment, even a 5% unemployment rate can be considered a serious political problem. Mill believes that the plural voting proposal will affect the quality of aims and values set by the citizens and the Parliament, and this will improve the quality of laws and policies since they will now be designed to achieve more valuable aims.



Mill's view is radically different from the thoughts of many contemporary political philosophers and epistemologists who discuss the role of experts in a democratic society. Philip Kitcher and Thomas Christiano, for example, agree that it is the role of a democratic process to set up important aims, and the role of experts to devise means for achieving these aims (Kitcher 2011, Christiano 2012). We should be democratic egalitarians when discussing political aims, and advocate expertism only when discussing practical means for achieving those aims. Mill disagrees and rejects democratic egalitarianism: there are those who are more competent in setting valuable aims and they should have greater political power in a democratic decision-making process. Of course, this does not imply that only those more competent should participate in the process of defining valuable aims, since that would reintroduce the danger of class legislation, but also damage the epistemic value of diverse perspective.

7. *Mill and public justification*

What makes plural voting procedure legitimate? As Estlund points out (Estlund 2003), Mill acknowledges the need for plural voting to be generally acceptable rather than simply correct. Authority does not follow from expertise, but from our acceptance that those wiser than us should have greater political power than us. This takes a form of hypothetical (or maybe normative) consent, and not a form of the actual consent.

It is only necessary that this superior influence should be assigned on grounds which [all] can comprehend, and of which [all] are able to perceive justice. (Mill 1977a: 474)

This is why Mill has to find a criterion for expertise that can be reasonably accepted by everyone. The problem is the fact that there is reasonable disagreement on who counts as wise. However, the idea that good education improves the ability to rule more wisely is uncontested.

[The distinctions in voting power] are not made arbitrary, but are such as can be understood and accepted by the general conscience and understanding. [They are based on something that] would not necessarily be repugnant to any one's sentiment of justice. (Mill 1977a: 476)

Finally, the reason why everyone should accept plural voting procedure is the quality of outcomes.

Which of these modes of getting over a difficulty is most for the interest of both, and most conformable to the general fitness of things? [...] that the better judgment should give way to the worse, or the worse to the better? (Mill 1977a: 473–474)

Since Mill believes that good education improves our ability to rule more wisely (i.e. to make better decisions), and since he believes that everyone shares (or should share) this belief, he emphasizes plural voting as a procedure that gives greater political power to those who can rule more wisely, and consequently favors it as a procedure that tends to create better outcomes.⁵

As we have seen, plural voting proposal has two goals: (i) to improve the quality of the outcomes by giving the educated additional political power, and (ii) to improve the intellectual and moral qualities in individuals by making them deliberate and exchange reasons and arguments.

8. Conclusion

Mill was undoubtedly one of the greatest liberal philosophers and an inspiring source of ideas for many liberal thinkers and scholars. We must, however, resist an increasingly common trend of interpreting Mill's ideas from the standpoint of contemporary liberal thought, especially when such interpretations contradict with the very statements Mill made himself. Plural voting proposal plays an important role in Mill's philosophical thought—it puts together and connects various requirements and values Mill held as important into a coherent collective decision-making model. This model does not rest on the idea of political equality, but on the complex structure that incorporates both the epistemic value of diverse perspectives and the epistemic value of experts. Consequently, it stresses both the importance of political participation and the impor-

⁵ One can consistently argue against this idea and rise against it not only argument based on procedural fairness, but an epistemic argument as well (see Estlund 2003).

tance of unequal political power citizens should have. We can discuss how Mill's ideas could be implemented in a contemporary liberal philosophy (e.g. is class legislation still an important issue, should everyone receive an equal chance of acquiring good education and thus greater political power, etc.), but we must not forget or misinterpret Mill's basic ideas and the justificatory process he made from them.

References

- Arneson, R. J. 2003. "Defending the Purely Instrumental Account of Democratic Legitimacy." *Journal of Political Philosophy* 11 (1): 122–132.
- Baccarini, E. 1993 *Sloboda, demokracija, pravednost. Filozofija politike J.S. Milla* (Freedom, Democracy, Justice: Political Philosophy of J.S. Mill). Rijeka: Hrvatski kulturni dom.
- Baccarini, E. 2013 "John Stuart Mill." In: Kulenović, E. (ed.) (2013) *Moderna politička teorija* (Modern Political Theory). Zagreb: Fakultet političkih znanosti.
- Baccarini, E. and Ivanković, V. 2015. "Mill's Case for Plural Voting and the Need for Balanced Public Decisions." *Prolegomena* 14 (2): 137–156.
- Berlin, I. 1969 *Four Essays on Liberty*. London: Oxford University Press.
- Brihante, A. A. and Rocha, F. J. 2013. "Democracy and Plural Voting in John Stuart Mill's Political Thought." *Ethic@* 12 (1): 53–65.
- Christiano, T. 2003. *Philosophy and Democracy: An Anthology*, Oxford: Oxford University Press.
- Christiano, T. 2004. "The Authority of Democracy." *Journal of Political Philosophy* 12 (3): 266–290.
- Christiano, T. 2008. *The Constitution of Equality*, Oxford: Oxford University Press.
- Christiano, T. 2012. "Rational deliberation among experts and citizens." In: Parkinson, J. (ed.) *Deliberative Systems: Deliberative Democracy at the Large Scale*. Cambridge: Cambridge University Press.
- Cohen, J. 1986. "An Epistemic Conception of Democracy." *Ethics* 97 (1): 26–38.
- Dahl, R. A. 1989. *Democracy and Its Critics*. London: Yale University Press.
- Estlund, D. 2003. "Why not Epistocracy?" In: Reshotko, N. (ed.), *Desire, Identity and Existence: Essays an Honor of T.M. Penner*. Kelowna: Academic Printing and Publishing.
- Estlund, D. 2008. *Democratic Authority*. Princeton: Princeton University Press.
- Gaus, G. 1996. *Justificatory Liberalism: An Essay on Epistemology and Political Theory*. New York: Oxford University Press.
- Gaus, G. 2008. "Is the public incompetent? Compared to whom? About what?" *Critical Review: A Journal of Politics and Society* 20 (3): 291–311.
- Honohan, I. 2002. *Civic Republicanism*. New York: Routledge.
- Irwin, T. H. 2006. "Mill and the Classical World." In: Skorupski, J. (ed.), *The Cambridge Companion to Mill*. Cambridge: Cambridge University Press.
- Justman, S. 1990. *The Hidden Text of Mill's Liberty*. Lanham: Rowman & Littlefield Publishers.

- Kitcher, P. 2011. *Science in a Democratic Society*. New York: Prometheus Books.
- Kristol, I. 1995. *Neoconservatism: The Autobiography of an Idea*. New York: Simon and Schuster.
- Macpherson, C. B. (2012) *The Life and Times of Liberal Democracy*. New York: Oxford University Press.
- Mill, J. S. 1879. *On Liberty*. New York: Henry Holt and Company.
- Mill, J. S. 1977a. *Considerations on Representative Government*. Toronto: University of Toronto Press.
- Mill, J. S. 1977b. *Thoughts on Parliamentary Reform*. Toronto: University of Toronto Press.
- Mill, J. S. 1978. *Essays on Philosophy and the Classics*. Toronto: University of Toronto Press.
- Miller, D. E. 2000. "John Stuart Mill's Civic Liberalism." *History of Political Thought* 21 (1): 88–113.
- Miller, J. J. 2003. "J. S. Mill on Plural Voting, Competence and Participation." *History of Political Thought* 24 (4): 647–667.
- Misak, C. 2000. *Truth, Morality, Politics: Pragmatism and Deliberation*. New York: Routledge.
- Peter, F. 2009. *Democratic Legitimacy*. New York: Routledge.
- Peter, F. 2012. "The procedural epistemic value of deliberation." *Synthese* 190 (7): 1253–1266.
- Peter, F. 2014. "Political Legitimacy." Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition).
- Pettit, P. 1999. *Republicanism: A Theory of Freedom and Government*. New York: Oxford University Press.
- Sandel, M. J. 2009. *Justice: What's the Right Thing to Do?* New York: Farrar, Straus and Giroux.
- Swift, A. 2006. *Political Philosophy*. London: Polity.
- Talisso, R. B. 2009. *Democracy and Moral Conflict*. New York: Cambridge University Press.
- Thompson, F. D. 1976. *John Stuart Mill and Representative Government*. New Jersey: Princeton University Press.
- Urbinati, N. 2002. *Mill on Democracy: From the Athenian Polis to Representative*. Chicago: University of Chicago Press.
- Wall, S. 2007. "Democracy and Equality." *The Philosophical Quarterly* 57 (228): 416–438.
- Young, I. M. 2000. *Inclusion and Democracy*. Oxford: Oxford University Press.

Is The Standard Definition of Knowledge Incomplete?

ANGUEL S. STEFANOV

*Institute for the Study of Societies and Knowledge
at the Bulgarian Academy of Sciences, Sofia, Bulgaria*

The aim of this paper is to suggest a new interpretation to the Gettier problem by showing that the standard JTB definition of knowledge is not epistemologically incomplete, being at the same time formally incomplete. The Gettier problem is shown to emerge through the implicit self-application of the JTB definition of knowledge to prove its own incompleteness. A conclusion is drawn, which runs counter to the traditional view that the problem necessarily requires a conceptual amendment of the standard definition, in spite of the formal incompleteness of the latter. The Gettier problem is construed to be related to the ancient Meno problem within a contemporary justificational discourse.

Keywords: JTB definition of knowledge, Gettier problem, Meno problem, justification, self-application.

1. Preliminary Remarks

The standard definition of knowledge was inspired by Plato's dialogues *Theaetetus*, *The Republic*, and mostly by *Meno*. To Meno's question "Why knowledge should be so much more prized than right opinion, and indeed how there is any difference between them" (*Meno* 97 d), Socrates explains his answer using a dialectical metaphor. It is based on the legendary masterhood of Daedalus to create statues in so perfect a manner that, if not tethered, they run away and escape:

Socrates: ...And that, I may say, has a bearing on the matter of true opinions. True opinions are a fine thing and do all sorts of good so long as they stay in their place, but they will not stay long. They run away from a man's mind; so they are not worth much until you tether them by working out the reason. (*Meno* 97e–98a)

The last metaphor of "tethering", or "working out the reason" of a true opinion (alēthēs dóxa), and thus turning it into knowledge (epistēmē),

was further elucidated by Socrates through his well known theory of recollection. Philosophers, who do not accept the literal message of Plato's epistemological view, have elaborated his approach, to reach the tacit agreement that knowledge is *justified true belief (JTB)*. And this standard JTB definition has seemed to be unproblematically indorsed (at least by analytic philosophers) until the beginning of the 60ies of the twentieth century.

Then in 1963 on the epistemological scene appeared Edmund Gettier. In a short, three pages article, he managed to destroy the plausibility of the accepted standard definition (Gettier 1963). He succeeded in so doing, by adducing two clear cases, showing that one could have some *true belief* that is also *justified*, and yet not have *knowledge*. So the claim has been raised that JTB is probably necessary, but not a sufficient condition for having knowledge, and that a better definition of knowledge must be discovered. Thus the notorious *Gettier problem* was born, and since then has not obtained a commonly accepted solution; or, as Robert Shope puts it:

In spite of the vast literature that Gettier's brief paper elicited, there is still no widespread agreement as to whether the Gettier problem has been solved, nor as to what constitutes the most promising line of research. (Shope 1998: 54)

A decade, or so, after this remark by R. Shope, "the literature on Gettier's brief paper" has rapidly enhanced. The standard strategy of searching for a definition of knowledge *stronger* than JTB, so that it could stay outside the pincers of the Gettier problem, has not resulted, however, in forging such a definition of knowledge to be generally accepted, or at least as tacitly accepted as JTB has been till 1963.

The aim of this paper is to suggest a new interpretation to the Gettier problem, by showing that its genesis lies in the implicit application of the JTB definition to gain its own incompleteness.

The Gettier problem, however, still persists in the renovated conceptual garments of the Meno problem about "tethering" mere true beliefs to be turned into knowledge. These garments bear the label of *justification*. So, it is the idea of justification that needs to be placed in the centre of a broader and deeper analysed.

2. *A New Look at the Gettier Problem*

After the publication of Gettier's paper, a good deal of Gettier-like counter-examples to the standard (JTB) definition of knowledge was suggested. Since they all have a common anti-luck-and-luck scenario, I'll prefer to turn back to the first original case, presented by Gettier himself.

Suppose that Smith and Jones have applied for a certain job. And suppose that Smith has strong evidence for the following conjunctive proposition:

d. Jones is the man who will get the job, and Jones has ten coins in his pocket.

Smith's evidence for (d) might be that the president of the company assured him that Jones would in the end be selected, and that he, Smith, had counted the coins in Jones's pocket ten minutes ago. Proposition (d) entails: e. The man who will get the job has ten coins in his pocket.

Let us suppose that Smith sees the entailment from (d) to (e), and accepts (e) on the grounds of (d), for which he has strong evidence. In this case, Smith is clearly justified in believing that (e) is true.

But imagine, further, that unknown to Smith, he himself, not Jones, will get the job. And, also, unknown to Smith, he himself has ten coins in his pocket. Proposition (e) is then true, though proposition (d), from which Smith inferred (e), is false. In our example, then, all of the following are true: (i) (e) is true, (ii) Smith believes that (e) is true, and (iii) Smith is justified in believing that (e) is true. But it is equally clear that Smith does not *know* that (e) is true; for (e) is true in virtue of the number of coins in Smith's pocket, while Smith does not know how many coins are in Smith's pocket, and bases his belief in (e) on a count of the coins in Jones's pocket, whom he falsely believes to be the man who will get the job (Gettier 1963: 121–2).

The key claim from the cited case above certainly is

(K) "Smith does not *know* that (e) is true".

Now let us pose the question: "Who is the *knower* of (K)?" The direct answer is that the knower is the teller of the story; let us accept that this is E. Gettier himself. But how he came to *know* that (K)?

This is a crucial question, because, if this claim were not true, then Gettier would not fulfil his task to show that JTB account of knowledge fails.

Of course, as a counterexample to a claimed theory of knowledge, it doesn't actually seem to matter whether anyone knows that the counterexample exists, or not. So, one could contend that Gettier is not supposed to prove that he knows that K. It is Smith who doesn't know that he himself will get the job, and that he has ten coins in his pocket, but nevertheless believes that (e) is true. Thus it seems that the knowledge of (K) is not related to the demonstration of the incompleteness of the JTB definition of knowledge.

However, it is the very definition of knowledge that is at stake here. To this effect Gettier (the teller of the story) must be certain of what he would like to convince us. It is no doubt that *an instrument for analysing knowledge (all the more its definition) has to be of a cognitive nature*; that is to say, it must not be *less* than knowledge, for instance a mere opinion, or a fabricated story, staying outside of some real cognitive context. Hence Gettier simply *has to know that (K)*. But in order for him to be a knowing agent – in this case a *meta-knowing agent* – he himself must be involved in a *real cognitive situation*, and not to be an inventive contriver of the adduced story.

So, how Gettier may really have knowledge that (K)?

The answer to this question would mean elucidating the fact that *he has a meta-knowledge in comparison to Smith*, who does not know that (e) is true. To this effect Gettier must be involved in a genuine situation

of *knowledge acquisition*.

Suppose that Gettier may “have strong evidence” that Jones has ten coins in his pocket, since he was a secret eye-witness when Smith was counting them “ten minutes ago”. He then is certain that Smith knows that Jones has ten coins in his pocket. Let us further suppose that he managed to count the coins in Smith’s pocket as well (when Smith was buying a cup of coffee at the near counter, for instance), and found that Smith had also ten coins in his pocket, and that he had also heard the words of the president of the company, when the latter was assuring Smith “that Jones would in the end be selected” for the job to which both men had applied. Then as the story goes, he correctly reaches the conclusion that (K).

But what does it mean that Gettier *has knowledge* that (K) in *his* situation of a meta-knowing agent? Let us pay attention to the fact that “all of the following are true”, to use Gettier’s own mode of argumentation: (iv) (K) is true, (v) Gettier believes that (K) is true, and (vi) Gettier is justified in believing that (K) is true.

Thus, as a meta-knowing agent, Gettier comes to know that (K) *in the way required by the JTB definition*. But it is namely the claim (K) that bears the burden of proving the insufficiency of *this same* definition of knowledge. So, what comes out is that the definition is being *implicitly* self-applied. We are facing a situation when a definition is appropriately used – an act presupposing its adequacy – to show its own inadequacy (its own incompleteness).

My new interpretation to the Gettier problem presented so far, does not show that the JTB definition of knowledge is a complete definition. It is, regrettably, formally incomplete. What I have, however shown, is that from an epistemological point of view this standard definition is not in need of an amendment, provided the conceptual requirements of its three partite structure are properly fulfilled.

3. *Conclusions*

The Gettier cases – as well as the first original case suggested by Gettier himself, and analyzed here – presuppose a concealed teller of a story suggesting a contrived situation, based on the JTB definition of knowledge, in which a knowing agent could be deceived that she knows something, while she does not. However, the very definition of knowledge that is here under attack lays the requirement that each fabricated story be considered within a real cognitive context. To this effect Gettier (the teller of the story) must be placed in a position of a meta-knowing agent. This is so, because an instrument for analysing knowledge (all the more its definition), a story no matter how fabricated, has to be of a cognitive nature. It was shown that at a meta-knowing level the JTB definition was kept intact. Thus *a specific negative answer to the title question was reached*: the standard (JTB) definition of knowledge is not *epistemologically* incomplete, being at the same time *formally* incom-

plete. The standard definition of knowledge does really work, provided the justification at hand is sound and trustworthy. If not, *it is not the JTB definition to be blamed*, but the cognitive potential of the knowing agent, and the contingent limitations of her interpretative context, so that a bad luck could not be evaded.

Having probably an intuition for a similar conclusion, Alvin Plantinga once wrote:

After 1963 the justified true belief account of knowledge was seen to be defective and lost its exalted status; but even those convinced by Gettier that justification (along with truth) isn't *sufficient* for knowledge still mostly think it *necessary* and *nearly* sufficient for knowledge: the basic shape or contours of the concept of knowledge is given by justified true belief, even if a quasi-technical fillip or addendum ("the fourth condition") is needed to appease Gettier. (Plantinga 1990: 45, his italics)

The here reached conclusion removes the necessity of "quasi-technical fillips" allegedly amending the JTB definition, enlarging it with "addenda". This "definitional" line of research, being otherwise stubbornly followed, may well be deserted. The Gettier problem is inspired by the ancient Meno problem within a contemporary justificational discourse. Instead of Plato's ancient metaphorical instruction of "tethering", or "working out the reason" of a merely true belief, the believer should care about its *adequate justification*.

By "adequate justification" I have in mind not the invention of some new *theory of justification* (that is certainly a theoretical ideal, but no less certainly a doubtful and problematic aim). I have in mind the specification of *paradigmatic justificational contexts* (e.g. mundane, juridical, scientific, etc.). For every cognitive situation in each of the paradigmatic contexts the explication of a relevant set of necessary criteria for justification is principally possible, so that the validity of stated propositions of interest to be warranted, and not merely guessed.

References

- Gettier, E. L. 1963. "Is Justified True Belief Knowledge?" *Analysis* 23: 121–123.
- Plantinga, A. 1990. "Justification in the 20th Century." *Philosophy and Phenomenological Research* 50: 45–71.
- Shope, Robert K. 1998. "Gettier Problems". In: E. Craig (ed.), *Routledge Encyclopedia of Philosophy, Vol.4*. London: Routledge.

Book Discussion

Consumer Complicity and Labor Exploitation

GILLIAN BROCK

University of Auckland, Auckland, New Zealand

Are consumers in high-income countries complicit in labor exploitation when they buy good produced in sweatshops? To focus attention we consider cases of labor exploitation such as those of exposing workers to very high risks of irreversible diseases, for instance, by failing to provide adequate safety equipment. If I purchase a product made under such conditions, what is my part in this exploitation? Is my contribution one of complicity that is blameworthy? If so, what ought I to do about such participation? I address these questions at first by applying a comprehensive account recently offered by Chiara Lepora and Robert Goodin, and analyzing the results in light of some important empirical issues.

Keywords: Consumer, complicity, labor exploitation, Lepora, Goodin, wrongdoing and exploitation.

1. Introduction

Many consumers in affluent developed countries are acutely aware of the horrendous levels of exploitation widespread in developing countries. Our iPhones are produced by exposing workers to toxic substances known to cause increased rates of leukemia. Our clothes are sewn by women working long hours, 7 days a week, in shoddily constructed buildings vulnerable to collapse. Children, who are forced by parents lacking income to work instead of attend school, weave the rugs and pillows that make our homes comfortable and aesthetically pleasing.

As a consumer in an affluent developed country, I and like-minded friends are often drawn to ask three questions about our prospective and actual purchases:

- 1) Are we complicit in the labor injustices that are inflicted on workers when we purchase such products?
- 2) If we are complicit, is this complicity morally culpable?
- 3) If so, what ought we to do to avoid or make amends for morally culpable complicity?

Chiara Lepora and Robert Goodin's recent book, *On Complicity and Compromise*, is an immensely helpful resource in addressing such questions (Lepora and Goodin: 2013). This work offers a comprehensive analysis of complicity and related concepts that significantly breaks new ground. This fine account is destined to shape conversations about complicity for some time and is well worth reading for anyone concerned about complicity in the many areas in which it seems to be playing an important normative role. In order to appreciate just some of the rich resources it offers I discuss some of its core ideas that have a bearing on the three featured questions outlined concerning consumer complicity with exploitation. Their book offers a wealth of other useful analysis especially concerning health care workers' complicity with genocide or torture, along with some excellent insights about choosing the least evil option when all the options one faces are bad. Here I focus only on applying their analysis to consumer complicity with exploitation.

2. *Exploitation and Wrongdoing*

For the concept of complicity to be in play, there must be some wrongdoing and actors must have played a part in bringing this about, sufficiently aware of these connections. Here I take the core wrongdoing to be that associated with exploitation in sweatshop labor conditions. Matt Zwolinski defines a sweatshop as "a place of employment in which worker compensation or safety is compromised, child labor is employed, and/or local labor regulations are routinely disregarded in a way that is prima facie morally objectionable".¹ While sweatshops exist all over the world, the focus of concern has been on sweatshops in developing countries. Three common kinds of problems with these sweatshops are often noted: (1) The wages are thought to be "objectionably low" (Zwolinski 2012: 162); (2) worker safety is seriously compromised by exposing employees to high risks of injuries or dangerous chemicals, without being provided with adequate safety training; (3) employers often suppress workers' rights to bargain collectively or unionize, or otherwise make severely unreasonable requests such as requiring employees to work long periods of overtime. The exploitation that occurs in sweatshop labor is a product of other, often prior, injustice. The kind of exploitation that concerns me involves taking advantage of people by

¹ As Matt Zwolinski continues, "This definition captures the role that the term 'sweatshop' plays as a signal of moral disapprobation, while leaving open as a conceptual matter the possibility that sweatshop practices might, on closer examination, be morally justifiable" (Zwolinski 2012: 154–179, 162).

taking advantage of their bargaining weaknesses. It is because people are in poor circumstances with few options for meeting basic needs that they typically find working in sweatshops attractive, compared with their alternatives.

Why does exploitative sweatshop labor count as wrongdoing? Accounts vary. Some argue that exploitation in sweatshops involves benefiting from others through unfairly taking advantage of them or their situation, benefiting from their misfortune or benefiting “disproportionately to their contribution” (Meyers 2004: 324). Or the emphasis could be on how the exploitative interactions can be degrading to those exploited, an affront to their dignity. According to an interesting recent analysis, Jeremy Snyder argues that when exploitative offers do not allow those exploited to make sufficient progress toward a decent minimum of human functioning, these offers can create “demeaning choices” (Snyder 2013: 346). The exploitee’s participation and apparent endorsement of the treatment, contributes to its being demeaning.

Generally, exploitation involves taking advantage of others in a way that is unfair or degrading and so appears to be squarely in the territory of wrongdoing. Some argue however that far from exploitative labor practices counting as wrongdoing they constitute right doing in that we are offering destitute people a way out of their situation. In fact, some argue that there can be forms of mutually voluntary and beneficial exploitation, where the exploitative transaction is beneficial to both parties, even when it is more beneficial to one than the other. Sweatshop labor is sometimes presented as exactly such a case (see e.g. Zwolinski 2012). In response, we might note that actions can be voluntary, beneficial and exploitative. The voluntariness of the exchange and benefits received do not necessarily cancel out or disguise aspects of a transaction that constitute wrongdoing, as wrongdoing may well remain. The very fact that these interactions are voluntary and mutually beneficial may serve to attempt to disguise that the moral wrong of exploitation has taken place. As I discuss in Section 4, a transaction can still be exploitative but all things considered, constitute the best course of action given undesirable alternatives.

However, we should also note that just how bad the wrongdoing is will vary depending on the nature and severity of the exploitation at issue. To focus our attention in this essay we might specify that the principal wrongdoings we are concerned with include cases of: (i) exposing workers to very high risks of irreversible diseases, for instance, by failing to provide adequate safety equipment; (ii) offering wages and/or terms of employment that thwart very basic needs (such as the need for rest or to take bathroom breaks); (iii) employing children in non-beneficial forms of labor that badly thwart their education or development. Here I am going to assume that there are at least some kinds of exploitation in sweatshop conditions that the reader would count as wrongdoing. If she is unconvinced by the examples provided she should

feel free to substitute her preferred examples for the three that I consider clear instances of objectionable wrongdoing.²

At any rate, I aim to address those readers who share my concerns about the kinds of examples described in my opening paragraph. I believe that there are enough consumers who view sweatshop labor exploitation as wrongdoing to make the analysis of interest. We want considered answers to core questions such as: What is my part in this wrongdoing? Is my contribution blameworthy? What ought I to do about my participation in wrongdoing? It is also important to note that the “we” I have in mind is composed of those consumers in high-income countries who have much disposable income with many consumption choices. They also have reasonable capacity to absorb costs because of their greater than average incomes within high-income countries.

3. *Complicity and its cousins*

There are many concepts that are nearby relatives of complicity. Some are much more straightforwardly problematic and easily implicate us in the wrongdoing, including full joint wrongdoing, conspiracy, and collusion. In such cases the culpability for the wrongdoing is easily locatable as there is a plan shared among the principal actors to commit the wrongdoing, to adopt the plan, or to orient behavior around a shared plan. In the cases that concern us, the consumers do not adopt a plan to exploit *as such*. Rather, their contribution to the wrongdoing is more nuanced and therefore more difficult to locate. But locate it we can.

3.1. *Complicity simpliciter*

Consider this passage from Lepora and Goodin:

As suggested by the loosely related legal concept of an ‘accessory’, those who are complicit simpliciter *often perform contributory acts that ‘give access’ to the principal wrongdoing, facilitating it or perhaps even making it possible*. Their contributions, although only ever causal (at most), may be more or less essential to the implementation of the principal wrongdoing. Or they might induce or incentivize the wrongdoing (a thief would not have stolen the painting *if there had not been anyone prepared to serve as the ‘fence’ in selling it*) or encourage it (a demagogic politician ranting against the excesses of banks encourages people to rob them) or make it easier to perform (selling robbers a precision drill, knowing the use they intend for it, halves the time it takes them to crack the safe). Temporally, acts of complicity simpliciter can come before, during, or after the principal wrongdoing.

Agents who are complicit simpliciter may act with more or less (but usually a minimal degree of) awareness of the details of the principal’s plan of wrongdoing. They might actually approve of the plan, or even participate in making it; they might adopt it as their own, and adjust their actions in response to it. But although full involvement in planning the wrongdoing is

² Perhaps these cases involve very coercive conditions; parents forcing their children into slavery-like conditions or prostitution might be of this kind. Sadly, this captures a staggeringly large number of cases in developing countries.

not excluded, *neither is that a necessary feature of an agent who is complicit simpliciter*. In order to qualify as complicit simpliciter, *all that is necessary is that the complicit agent 'knows, or should have known, that by [so acting] he or she will advance whatever intentions the principal has'*. (Lepora and Goodin 2013: 42, emphasis mine)

On reading this passage a few comments might immediately strike us as worth mentioning. First, typical sweatshop labor consumption acts seem relevantly analogous with the stolen painting case. If no one is willing to purchase the products made with sweatshop labor it would not be worth the manufacturer's efforts to engage in these production processes. But manufacturers can be reasonably sure there will be willing consumers because, for many consumers, the price is right. Second, we might want to know more about what is entailed by the phrase "knows or should have known". What do reasonable consumers know or what should such consumers know? There is considerable variation about cases. For instance, when a product is marked "Made in the USA" although every component part is made elsewhere under sweatshop conditions, should the consumer suspect deception? How much due diligence should she show in trying to check marketer's claims about their products? What does the reasonable consumer know and what can we reasonably expect of her in investigating the labor conditions under which all her prospective purchases are made? While these questions constitute an important area ripe for extended analysis, here I simplify by making some assumptions. I assume reasonable consumers are generally informed about the world they inhabit and that they can be expected to have general knowledge of working conditions in a globalized world. So they can be expected to know that many of the products they purchase from stores like Walmart are the product of sweatshop labor.

3.2. *Complicity by collaboration*

When we buy goods made in sweatshops are we collaborating with producers? Collaborators are not co-principals of the wrongdoing. In fact, they need not even adopt plans to commit wrongdoing.

The relationship between the collaborator and the principal is purely that of follower to leader, in regard to the plan. The collaborator takes instructions from the plan and adjusts his own actions to it. Collaboration involves the active and practical engagement of a contributory agent with a plan *that in some way he accepts and acts upon*. But while accepting the plan as a basis for his actions, *he need not actually adopt the plan as his own*. The collaborator's stance toward the plan might be far more equivocal than that. (Lepora and Goodin 2013: 43)

This passage contains an important tension. On the one hand, you might think complicity by collaboration is not relevant as there is no real acceptance of the plan to commit wrongdoing when consumers typically purchase sweatshop labor products. On the other hand, since Lepora and Goodin stipulate that the co-principal need not adopt the

plan as his own, maybe this is not so clear. At any rate, in eschewing the exploitation part of the plan, perhaps thoughtful consumers are not guilty of complicity by collaboration, when they purchase certain kinds of products that result from exploitative labor practices. As the other forms of complicity to be discussed provide much clearer cases of implication in wrongdoing, we need not dwell on complicity by collaboration, which might at most be an atypical case.

3.3. *Complicit acts that (can misleadingly) appear not to involve contributors*

According to Lepora and Goodin, complicity “necessarily involves acting in a way that *could contribute causally to the principal wrongdoing of another*” (Lepora and Goodin 2013: 44). While at first it might seem that this cannot apply to what they refer to as “the conceptual cousins” that will be our focus here, namely, conniving, condoning, consorting, or contiguity, this appearance is misleading. In general there can be ways in which actions in response to wrongdoing promote future wrongdoing, so consumers’ actions can have a causal role in wrongdoing after all. I discuss some of the ways in which “complicity’s cousins” can implicate many consumption acts.³

3.3.1. *Condoning*

While condoning isolated wrongs could be wrong, it cannot be so in a *causal* way, since the wrongdoing has already taken place. However, an act of condoning can qualify as a relevantly causal contribution under certain conditions. Such special cases are called “complicity by condoning”. They say: “While you cannot literally ‘condone’ a wrong ahead of it occurring, you can announce ahead of time that you will condone it were it to occur; and that announcement (while itself not literally an act of condoning) can certainly contribute causally to the act-to-be condoned occurring” (Lepora and Goodin 2013: 48). Perhaps consumers’ purchasing patterns could count as signaling ahead of time that they will condone exploitative production acts, at least in the sense that they could be construed as announcing that they will purchase the products of sweatshop labor anyhow despite the wrongdoing. Heavily repeated patterns of consumption or strong consumer demand can count as a kind of condoning as consumers thereby seem to be signaling that they accept an ongoing practice of wrongdoing which can contribute in a causal way to subsequent exploitation.

3.3.2. *Consorting*

Complicity by consorting can make a causal contribution to the principal wrongdoing, when “consorting with the wrongdoers, signals one’s

³ These are discussed in likely descending order of normative importance.

agreement with and approval of their actions; and that encourages them in their wrongdoings” (Lepora and Goodin 2013: 49). As with complicity by condoning, ongoing patterns of consumption could certainly serve this signaling function. By patronizing and continuing to purchase from suppliers known to perpetrate exploitation, we could reasonably be accused of complicity by consorting.

3.3.3. *Contiguity*

Complicity by contiguity seems quite relevant as well. When principal wrongdoers interpret “the secondary agent’s contiguity as *implicit approval* of a wrong, she might be encouraged in her wrongdoing” (Lepora and Goodin 2013: 51). If the secondary agents know (or could and should have known) that their contiguity would be interpreted in such a manner or could have such effects, then their “continuing voluntary contiguity to wrongdoing might constitute complicity with the wrongdoing” (Lepora and Goodin 2013: 51). So consumers who repeatedly purchase from the same sweatshop supplier would seem to be complicit by contiguity. Repeatedly purchasing products or showing brand loyalty probably are sufficient to count as complicity by contiguity because reasonable consumers should know that this could be taken as signaling sufficient approval.

3.3.4. *Connivance*

Connivers’s participation can range from ignoring another’s wrongdoing to being in secret sympathy or even encouraging the wrongdoing. While connivers do not participate in making the plan, nor adopting it, they “stand aside to allow others to act on it” (Lepora and Goodin 2013: 44). Cases of allowing repeated patterns of wrongdoing can count as connivance when similar situations involving the same agents recur regularly, since “acts of connivance with a wrong committed today may contribute causally to the wrongdoer’s repeating the wrong on the next occasion. It does so by making the wrongdoer confident, on the basis of previous experience, that again in the future onlookers will connive rather than intervening to stop the wrong when they see it occurring” (Lepora and Goodin 2013: 46). Perhaps consumers can then be complicit by connivance when they allow exploitative labor to continue and fail to take relevant actions in response to labor injustices. In so far as they “stand aside” in these ways and fail to take actions to end the wrongdoing, they are guilty of complicity by connivance.

3.4. *Examples*

Lepora and Goodin discuss a few salient examples which illustrate how a variety of forms of complicity may occur together, such as that of Oskar Schindler who despite early complicity with the Nazi regime in manufacturing armaments, famously also employed more than 1200

Jewish workers in his factory who were otherwise to be deported to concentration camps and face almost certain death. The other example involves bank robbers enlisting the help of a taxi driver to make a get-away, even though the driver was not involved in the bank robbery itself. Both the post-robbery taxi driver and the Schindler case are quite relevant as you might think the central problems with exploitation are characterizable as either ones of essentially stealing (not compensating workers fairly for their labor) or of saving them from a worse fate (death or further deprivation).

According to Lepora and Goodin in “willingly and knowingly driving away robbers with money that he knows they have stolen in exchange for a share of it” (Lepora and Goodin 2013: 56) the taxi driver enlisted after the robbery can rightly be accused of collaborating with the robbers. Arguably, consumers in developed countries, seem to be collaborating with the manufacturers to some extent because they willingly and knowingly purchase the goods mindful that they are a result of exploitative labor practices. We might reasonably argue (as the authors do) that what Schindler did was right and what the taxi driver does is wrong. At any rate, even when we are complicit it is a further question how to evaluate our act of complicity, which brings us to our core normative idea for analysis.

4. Just how bad is it to be complicit in exploitation?

There are several different ways of engaging with the wrongdoing and some make more of a contribution to it. Some are much worse than others. What makes something more of a contribution to wrongdoing? There are two different dimensions: One involves the agent’s role in doing or contributing to the wrongdoing. The second relates to the agent’s mental stance towards the main wrongdoing and towards the plan of action related to it.

For morality to be action guiding, it should “assess the situation in terms of what can be known by the actor at the time she acts” (Lepora and Goodin 2013: 61). So when we conduct moral assessments of actions we should do so on the basis of what consequences might reasonably have been expected at the time of the decisions concerning action, not on the basis of what actually happened. Consumers might have various beliefs about the consequences that can reasonably be expected from their purchases and these beliefs can be relevant to moral assessment of their actions. In particular, consumers might have relevant beliefs about the longer-term consequences of the exploitation, even if they believe that the current exploitation constitutes wrongdoing. Consider two such beliefs:

- 1) Exploitation leads to the long-term promotion of better conditions for workers.

- 2) Accepting an offer to work in exploitative labor conditions is better than having no source of income at all (and no other way to meet one's basic needs).

What would believing (1) or (2) (or both) mean for a moral assessment of our purchase? In order to address such issues we need to discuss briefly the plausibility of such views (which we do in Sections 4.1 and 4.2). Then in Section 4.3 I argue that even if the two views are plausible, we might still be culpable for our part in current wrongdoing, even when participating in current wrongdoing is the best course of action available to us.

4.1. *Does exploitation lead to the long-term promotion of better conditions for workers?*

Economists often observe that exploitative labor practices in sweatshops exist not because there are greedy employers who cut corners for extra profits, but rather for more fundamental reasons. Consider how in Asia, the country wage rate mean is about 44c per hour of labor and in Latin America \$1.34. Why would those in Asia work for much less than those in Latin America? On a standard economic account:

...wages fall within a range limited by the marginal productivity of labor at the high end and the wages offered by the next best alternative available to workers at the low end. Employers will not pay workers more than what the last worker hired contributes to the firm's revenue, and employees will not accept wages lower than they could secure in alternative employment. The actual wage rate paid within that range depends upon the relative bargaining strength of the two sides.

The employment alternatives available to workers are in large part determined by their country's level of economic development. ... Unlike economically developed countries where the percentage of the labor force employed in agriculture typically is around 5 percent, in [developing] countries that percentage is much higher. For the Asian countries, the labor force in agriculture ranges from China's 72 percent to Sri Lanka's 35 percent.

So, the clothing industry (to take an example often implicated in sweatshop labor) has to offer only a slightly higher wage than the one paid in agriculture to attract workers. Garment makers work for much less in Asia mainly because there is an absence of high-paying alternatives that would come from more economic development. The lack of union strength in Asia is also an important factor in why wages there are so low.

One important point that is relevant from this brief excursion into economic theory is that the level of development in the country makes an important contribution to increasing options and reducing exploitative ones. To the extent that sweatshops bring capital, technology, opportunities for knowledge transfer, and so forth to developing countries, they can thereby raise, over time, worker productivity, income,

and growth (and can create other positive externalities), and all of this can contribute to beneficial development over the longer run. If that is correct – and the history of developed countries suggests it well might be – sweatshops might be a necessary evil along the way to the very kinds of better jobs we hope to secure for the world’s most disadvantaged. So, those who engage in sweatshop labor create a better future (in due course) for others. Those who take up such work provide valuable contributions to securing a better economic future for fellow citizens. Perhaps their heroic self-sacrificing actions should be supported.

4.2. Accepting an offer to work in exploitative labor conditions is better than having no source of income at all (and no other way to meet one’s basic needs).

A consumer might have the view that, all things considered, laborers should accept exploitative work when it is better than more dire alternatives, so we should buy goods produced in sweatshops. This casts doubts on whether the exploitation aspect of production is really bad. I am not convinced by such arguments. Consider an analogy closer to our academic experiences, which I call *Teaching Assistant Contract*. Many Teaching Assistants are not adequately remunerated for the number of hours they actually spend on grading and writing comments on students’ essays. As one example, consider contracts that assume graders take no more than 20 minutes to grade each essay when standardly the work involves at least 40 minutes. I can recognize that this component of their contract is exploitative (in the sense that they are not adequately remunerated for their labor) even if, on balance, I think it is better for graduate students to accept these exploitative contracts than not. It still may be better for them to accept contracts at a marginally higher rate of pay doing work that makes better use of their skills and talents than to accept work in (say) a fast food restaurant. So I think we can, do and should make these distinctions about how exploitation is descriptively accurate even when, all things considered, an exploitative contract ought to be accepted.

So, even if we believe the kinds of reasons offered under 4.1 and 4.2—that accepting exploitative work may lead to good consequences or is best given other options—we are still contributing to exploitative acts here and now. These two kinds of relevant beliefs are insufficient to cancel out complicity with exploitation. The real question is about how bad such complicity is.

4.3. Assessing Acts of Complicity: A General Framework

Lepora and Goodin offer a useful formula which enables moral assessment of how blameworthy complicit acts are. The moral assessment

is a function of four things; the moral badness of the principal wrongdoing; whether (and, insofar as it is scalar, by how much) the secondary agent

crosses the threshold of moral responsibility for having contributed to it; how much of a contribution his act made (or might make) to the principal wrongdoing; and the extent to which the secondary agent shares the purposes of the principal wrongdoer. Phrased as a formula:

Pro tanto blameworthiness for an act of complicity = function of (badness of principal wrongdoing, responsibility for contributory act, extent of contribution, extent of shared purpose with principal wrongdoer). (Lepora and Goodin 2013: 98)

A complete assessment of blameworthiness of a particular contributory act also requires assessment of the alternative courses of action. Only such a comprehensive analysis will allow an overall assessment of agent's choices.

Let us then apply this formula to our core case of concern. First, the badness factor. How bad is exploitation? While this can vary a whole lot, in the cases that we are taking as central, it is rather bad. Compare *Teaching Assistant Contract* with working in unsafe conditions that subject one to irreversible debilitating disease. We can make reasonable judgments that exploitation in the latter case is worse than that in the former, since it affects more significant interests and, in eventually affecting quality of life and shortening it, we can judge this as a worse wrongdoing than the wrongdoing involved in *Teaching Assistant Contract*.

Responsibility is determined by a combination of voluntariness, knowledge of contribution, and knowledge of wrongness of principal wrongdoing. For the cases I am taking as typical here, consumers perform contributory acts voluntarily. For an enormous range of cases, such consumers have many consumption options, though interestingly, this may not apply to the purchase of certain electronic goods like cell-phones and laptops, which require the use of hazardous chemicals in the production process. (Of course, even if it is the case that these goods cannot be produced without exposure to hazardous chemicals such as benzene, the safety training and protective equipment offered could be vastly improved over current prevalent levels.) I have suggested that a reasonable consumer should be in possession of general facts about supply chain employment conditions in a globalized world so she should know how she is contributing. And one should be aware that exploitation is wrong, even if all things considered it is permissible in certain cases.

Evaluating contribution to wrongdoing is a bit more difficult. It approximates to "the percentage of badness of that principal wrong that might be causally attributable to the contributory agent, by virtue of her contributory act" (Lepora and Goodin 2013: 106). Contribution is a factor combining a number of variables such as centrality of contribution, reversibility of contribution, temporality, planning role, and responsiveness of contributors to principals. Any particular consumer's consumption choice is not essential or central to the wrongdoing.

Whether or not the contribution is irreversible varies depending on the kind of exploitation. With exposure to toxic chemicals or denying a child proper education, the effects are frequently irreversible. The wrongdoing might be part of an ongoing pattern of similar wrongdoing. So the wrongs may be repeated more-or-less frequently. Consumers are not typically involved in planning the exploitation and do not share the sweatshop owners' purposes. They would prefer to avoid entanglement with exploitation if they easily and costlessly can.⁴

So where does that leave us? It seems clear that our purchases frequently involve morally culpable complicity, especially when workers are exposed to high risks of irreversible damage. What should we do in the face of morally culpable complicity?

5. *Morally culpable complicity: now what?*

Using the comprehensive analysis offered, I have argued that developed world consumers are often culpably complicit in exploitative labor conditions in developing countries. What follows from this verdict? It is not part of the Lepora and Goodin analysis to treat such questions, but let us anyhow consider some possible ways in which we might want to use their account to discuss these issues.

In the case of consumers purchasing products made involving exploitative labor practices there is no straightforward recommendation of what such consumers ought to do to mitigate future culpable complicity or remedy past complicity. We would need to rely on detailed accounts about what courses of action would be effective in reducing exploitation and what remedial actions might be appropriate. Elsewhere I argue that because exploitation in the cases that concern us involve taking advantage of others by taking advantage of their bargaining weaknesses, our aim should be to remedy that background situation that allows these bargaining weaknesses to continue (Brock 2014). There I have suggested we can and should do this in multiple ways, such as by strengthening collective organizational capacity, promoting effective and legitimate states along with active citizenship. There is plenty we can and should do *as citizens* implicated in the poor bargaining positions of those who feel compelled to accept exploitative offers.

But does that mean we have nothing to say specifically about *consumer* responsibilities here and now for limiting culpable complicity? There are things consumers ought to do in their role specifically as

⁴ As they move from considering *pro tanto* to on-balance judgments they discuss a number of useful cases such as the Nazi postman, someone who simply delivers mail for the Nazis. His contribution to wrongdoing is low and the probability that his contribution is essential to wrongdoing is tiny. However, "contribution factor has to be multiplied by the badness factor in assessing the overall *pro tanto* blameworthiness of the postman's contributory acts. And since the badness of the principal wrongdoing, the Holocaust, is so very large, the overall *pro tanto* blameworthiness of the postman for his contribution to the wrongdoing might be quite high, his low contribution factor notwithstanding" (Lepora and Goodin 2013: 119).

consumers (rather than citizens) and these include supporting certain kinds of efforts that are being made to empower and promote the condition of vulnerable workers, including supporting ethical or fair trading initiatives (importantly, those that ensure the gains and risks of trade are fairly distributed, along with respecting people's basic human entitlements) (see Brock 2014). These responsibilities are particularly salient for those I identify as the target consumers of concern. They satisfy a number of salient conditions that mean they have enhanced responsibilities (including having high capacity to assist, being relatively privileged, have benefited greatly from the exploitative practices, and so on).

However, as I have been discussing, there are not always ethical or fair trade options available, especially in the purchase of certain electronic equipment (such as cell phones and iPods). Here drawing our attention to our culpable complicity is nevertheless helpful in building an awareness of the many ways in which living in affluent countries culpably implicates us in practices that perpetuate harm. Here perhaps we are all a bit like Oskar Schindler in that we must choose the lesser evil, though in this case the lesser evil might entail purchasing products that provide the destitute with jobs even while they inflict high risk of irreversible disease. Of course, we have more options than Schindler did concerning how to affect change so the background conditions of severe deprivation do not continue and also so that protections against irreversible effects can be improved.⁵ Mindful of our culpable complicity we should be energized to learn more about how we can and should assist in reducing our complicity in harmful exploitation through our consumption.

References

- Brock, G. 2014. "Global Poverty, Decent Work, and Remedial Responsibilities: What the Developed World Owes to the Developing World and Why." In D. Meyers (ed.) *Poverty, Coercion, and Human Rights*, New York: Oxford University Press: 119–145.
- Lepora, C. and Goodin, R. 2013. *On Complicity and Compromise*. Oxford: Oxford University Press.
- Meyers, C. 2004. "Wrongful beneficence: Exploitation and third world sweatshops." *Journal of Social Philosophy* 38: 319–333.
- Zwolinski, M. 2012. "Structural Exploitation." *Social Philosophy and Policy* 29: 154–179.
- Snyder, J. 2013. "Exploitation and demeaning choices." *Politics, Philosophy and Economics* 12: 345–360.

⁵ The situation of Schindler's complicity with the Nazis is not a good model for the range of options available to us today. We are reasonably free to embark on a number of courses of action aimed at eliminating or at least reducing the forms of exploitative labor we find so rampant in the world today. What exactly is the most effective course of action available to us that could end these abuses is an enormous issue and I begin the analysis in works such as Brock 2014.

Croatian Journal of Philosophy is published three times a year. It publishes original scientific papers in the field of philosophy.

Croatian Journal of Philosophy is rated INT2
(in the field of Philosophy) and INT2
(in the field of Philosophy of Science) in the
European Reference Index in the Humanities (ERIH).
It is indexed in *The Philosopher's Index*
and in *Arts & Humanities Citation Index (Web of Science)*.

Payment may be made by bank transfer
SWIFT PBZGHR2X
IBAN HR4723400091100096268

Instructions for Contributors

All submissions should be sent to the e-mail: cjp@ifzg.hr. Submissions must be in English and formatted to be double-spaced with suitably wide margins, an A4 page size, and automatic page numbering.

Articles are normally no longer than 8,000 words of main text including bibliography. The Journal will consider longer papers, but, once these limits are exceeded, authors should bear in mind the editorial policy that the acceptance bar raises with increasing length.

Manuscripts should be compiled in the following order: cover page; title; abstract (not exceeding 200 words); keywords (3 to 6); main text; appendices (as appropriate); references.

All the authors of a paper should include their full names, affiliations, postal addresses, telephone and fax numbers and email addresses on the cover page of the manuscript. If a paper is co-written, one author should be identified as the Corresponding Author. The cover page must be submitted as a separate document. All submitted manuscripts must be prepared for blind review, with revealing acknowledgements and self-identifying references removed.

Sources are cited in the text by the author's last name, the publication date of the work cited, and a page number if needed, e.g. (Barber 2007: 324). Full details appear in the reference list in which the year of publication appears immediately after the author's name:

Barber, Alex. 2007. "Linguistic Structure and the Brain." *Croatian Journal of Philosophy* 21 (7): 317–341.

Williamson, Timothy. 2013. *Identity and Discrimination*. Oxford: Wiley-Blackwell.

ISSN 1333-1108



9 771333 110001