CROATIAN

JOURNAL

OF PHILOSOPHY

CROATIAN

JOURNAL

OF PHILOSOPHY

Vol. XVII · No. 51 · 2017

Book Reviews

On Stephen Neale's manuscript
Silent Reference

Introduction

Dubrovnik conference on the philosophy of language and linguistics which took place September 12 to 16, 2016 at the IUC continued with still growing interest in the "semantics-pragmatics" distinction, this time partly focusing on Stephen Neale's latest manuscript "Silent Reference". The first four papers are reactions to Neale's work, which will, hopefully, be followed by Neale's replies to his critics in the first issue of the Croatian Journal of Philosophy *in 2018.*

Stephen Schiffer's *paper "Gricean Semantics and Vague Speaker-Meaning" argues that presentations of Gricean semantics, including Stephen Neale's in "Silent Reference," totally ignore vagueness, even though virtually every utterance is vague. Schiffer asks how Gricean semantics might be adjusted to accommodate vague speaker-meaning. His answer is that it can't accommodate it: the Gricean program collapses in the face of vague speaker-meaning. The Gricean might, however, find some solace in knowing that every other extant meta-semantic and semantic program is in the same boat.*

Daniel W. Harris *in his contribution "Speaker Reference and Cognitive Architecture" points out that philosophers of language inspired by Grice have long sought to show how facts about reference boil down to facts about speakers' communicative intentions. He focuses on a recent attempt by Stephen Neale who argues that referring with an expression requires having a special kind of communicative intention—one that involves representing an occurrence of the expression as standing in some particular relation to its referent. Neale raises a problem for this account: because some referring expressions are unpronounced, most language users don't realize they exist, and so seemingly don't have intentions about them. Neale suggests that we might solve this problem by supposing that speakers have nonconscious or "tacit" intentions. Harris argues that this solution can't work by arguing that our representations of unpronounced bits of language all occur within a modular component of the mind, and so we can't have intentions about them. From this line of thought, Harris draws several interesting conclusions*

In his paper *"Saying without Knowing What or How"* Elmar Unnsteinsson *in response to Stephen Neale, argues that aphonic expressions, such as PRO, are intentionally uttered by normal speakers of*

natural language, either by acts of omitting to say something explicitly, or by acts of giving phonetic realization to aphonics. He also argues that Gricean intention-based semantics should seek divorce from Cartesian assumptions of transparent access to propositional attitudes and, consequently, that Stephen Schiffer's so-called meaning-intention problem is not powerful enough to banish alleged cases of over-intellectualization in contemporary philosophy of language and mind.

Jesse Rappaport *in his paper* "Is There a Meaning-Intention Problem?" *attempts to articulate the assumptions that support the meaning-intention problem. He argues that these assumptions are incompatible with some basic linguistic data. For instance, a speaker could have used a sentence like "The book weighs five pounds" to mean that the book weighs five pounds on Earth, even before anyone knew that weight was a relativized property. The existence of such "extrinsic parameters" undermines the force of the meaning-intention problem. However, since the meaning-intention problem arises naturally from a Gricean view of speaker's meaning and speaker's reference, the failure of the argument raises problems for the Gricean. He also argues that the analysis of referring-with offered by Schiffer, and defended by Neale, is defective.*

Two more articles conclude this issue. Erich Rast *in his paper* "Value Disagreement and Two Aspects of Meaning" *discusses two attempts of solving the problem of value disagreement: contextualist, relativist and metalinguistic. Although the metalinguistic account seems to be on the right track, it is argued that it does not sufficiently explain why and how disagreements about the meaning of evaluative terms are based on and can be decided by appeal to existing social practices. As a remedy, it is argued that original suggestions from Putnam's "The Meaning of 'Meaning'" ought to be taken seriously. The resulting dual aspect theory of meaning can explain value disagreement in much the same way as it deals with disagreement about general terms. However, the account goes beyond Putnam's by not just defending a version of social externalism, but also defending the thesis that the truth conditional meaning of many evaluative terms is not fixed by experts either and is instead constantly contested as part of a normal function of language.*

Mark Steen *in* "Temporally Restricted Composition" *develops and defends a novel answer to Peter van Inwagen's 'Special Composition Question,' (SCQ) namely, under what conditions do some things compose an object? His answer is that things will compose an object when and only when they exist simultaneously relative to a reference frame (He calls this 'Temporally Restricted Composition' or TREC). He then shows how this view wards off objections given to 'Unrestricted Mereology' (UM). TREC, unlike other theories of Restricted Composition, does not fall prey to worries about vagueness, anthropocentrism, or arbitrariness. TREC also has other advantages.*

DUNJA JUTRONIĆ

Gricean Semantics and Vague Speaker-Meaning

STEPHEN SCHIFFER
New York University, New York, USA

Presentations of Gricean semantics, including Stephen Neale's in "Silent Reference," totally ignore vagueness, even though virtually every utterance is vague. I ask how Gricean semantics might be adjusted to accommodate vague speaker-meaning. My answer is that it can't accommodate it: the Gricean program collapses in the face of vague speaker-meaning. The Gricean might, however, find some solace in knowing that every other extant meta-semantic and semantic program is in the same boat.

Keywords: Gricean semantics, Stephen Neale, proposition, vagueness, vague speaker-meaning.

Gricean semantics is a program for reducing all questions about the intentionality of speech acts and linguistic expressions to questions about the intentionality of thought. It takes as its cornerstone a certain notion of speaker-meaning and seeks to define it, without recourse to any semantic notions, in terms of acting with certain audience-directed intentions. Then it seeks to define other agent-semantic notions—most notably speaker-reference and illocutionary act—in terms of its defined notion of speaker-meaning, and after that to define the semantic features of linguistic expressions—meaning, reference, truth conditions, etc.—wholly in terms of the already defined agent-semantic notions, together with certain ancillary notions, such as that of convention, which are themselves explicable wholly in terms of non-semantic propositional attitudes. Gricean semantics is the program that Stephen Neale sympathetically explains with marvelous clarity and comprehensiveness in "Silent Reference."

Neale is a self-proclaimed Gricean. At the same time, he is aware that the Gricean program is not without problems. One of those problems is the possibility of "aphonics," indexicals that appear in a sentence's LF representation but are "silent" in that they aren't phonetically realized. For if there are aphonics, then they will evidently have referents rela-

tive to utterances of sentences in whose LF representations they occur, but it's difficult to see how those referents can be determined by speakers' referential intentions, which is what a Gricean account of expression-reference would seem to require. Neale would like to remove the threat posed by aphonics by showing that they can be accommodated within a Gricean theory of reference, and he suggests one way that accommodation might be achieved. But, as I tried to show (Schiffer 2017) in my response to "Silent Reference," there is a question as to whether that attempt succeeds, and I suggested another route by which accommodation might be achieved. This essay, however, isn't about aphonics. It's about a problem for which I believe no Gricean solution is possible.

A striking feature of presentations of Gricean theories of meaning—as well as of virtually every other presentation of a foundational theory of meaning—is that they completely ignore vagueness, even though virtually every utterance is vague. Perhaps the authors of these presentations would say that their ignoring vagueness is a useful idealization akin to Galileo's ignoring friction in his idealized model of bodies in motion. They might say that. But they would be wrong: vagueness can't be accommodated within the Gricean program. There is more than one way to show this, but I will begin by showing that no Gricean account of speaker-meaning can accommodate vague speaker-meaning, and that will be enough to enable us to see how vagueness also frustrates the aspirations of Gricean accounts of expression-meaning. We will also have an indication of how the problem that vagueness makes for Gricean semantics is one it also makes for competing semantic programs.

1. *Gricean Speaker-Meaning and Speaker-Reference*

Assertoric speaker-meaning—henceforth simply speaker-meaning—is the notion of a speaker's meaning that such-and-such, as when, for example, in uttering 'He's ready' Jill meant that Jack was ready to go to dinner. It's the most general kind of assertoric illocutionary act, the genus of which all other kinds of assertoric illocutionary acts—saying that such-and-such, asserting that such-and-such, denying that such-and-such, objecting that such-and-such, telling so-and-so that such-and-such, etc.—are species. The Gricean, like most current philosophers of language, takes speaker-meaning to be a relation, *S meant p*, between a person S and a proposition p that she meant, where a proposition is an abstract entity that has truth conditions, has those truth conditions necessarily, and is mind- and language-independent in that it belongs to no language and wasn't created by what anyone said or thought.[1]

---

[1] For my purposes it doesn't matter to which kind of proposition—Fregean, Russellian, functions from possible worlds into truth-values, whatever—the Gricean thinks the propositions we mean belong, but for simplicity of exposition I will often write as though the Gricean takes the propositions we mean to be Russellian propositions, i.e. structured entities whose basic components are the objects and properties our speech acts are about.

What distinguishes the Gricean from other theorists who take speaker-meaning to be a relation to propositions is the Gricean's conception of the nature of that relation: for the Gricean the speaker-meaning relation is definable wholly in terms of a speaker's intentions. Neale gives the flavor of a Gricean account of speaker-meaning with the following slightly altered capsule restatement of the account of speaker-meaning I proposed in (1972) and then later used to devise the concomitant accounts of speaker- and expression-reference Neale discusses in "Silent Reference":

> In uttering x, S meant p ≈ for some audience A and feature φ, S uttered x intending S and A to mutually know that x has φ and, on the basis of this, that S uttered x intending A actively to believe p.

Here 'utter' is used, following Grice (1957), in an artificially extended way to include any action by which a person might mean something—spoken utterances, of course, orthographic inscriptions, hand signals, as well as nonconventional acts by which a person might mean something, as when, for example, a person mimes being asleep to communicate that she is bored. In (what is supposed to be) the normal case in which a person means a proposition, she does so by uttering a sentence with whose meaning the proposition she means comports (as, for example, the proposition that it's raining in Dubrovnik comports with the meaning of the sentence 'It's raining') but for the Gricean it's essential that his account of speaker-meaning not in any sense require x to have antecedent meaning, for a defining feature of his theory of meaning is that expression-meaning is to be reductively defined in terms of his definition of speaker-meaning. 'Mutual knowledge' holds a place for whatever is the correct way of spelling out the requirement that S's meaning-constituting intentions be "out in the open" between her and her audience; David Lewis's notion of common knowledge (1969: 1975) and my similar notion of mutual knowledge* (1972) were independent attempts to do that spelling out, but for present purposes no specific account of mutual knowledge needs to be presupposed. An activated belief is a belief that one has consciously in mind. The most common way for S to intend A actively to believe p is for her to intend her utterance to cause A to believe p, as when the purpose of S's utterance is to provide A with knowledge of p; but S's intention in uttering x may be merely to remind A of p, and that, too, counts as S's intending A actively to believe p.

Corresponding to the distinction between speaker-meaning and expression-meaning is a distinction, central to the main issue joined in "Silent Reference," between speaker-reference and expression-reference. But this second distinction folds into the first, since, as Neale makes clear, the primary notion of speaker-reference is simply a species of speaker-meaning. Actually, there are three notions of speaker-reference to be accommodated, which I will call *primary speaker-reference*, *higher-order speaker-reference*, and *referring-with*. Primary

speaker-reference—signified by the subscript 'π' in 'S referred$_\Pi$ to o in uttering x'—is for the Gricean the most basic notion of speaker-reference in terms of which the other two notions are to be defined. It's defined thus:

> S referred$_\Pi$ to o in uttering x iff for some o-dependent proposition p$_o$, S meant p$_o$ in uttering x.

An o-dependent proposition is a proposition that is individuated partly in terms of o, so that o occurs essentially in the proposition's truth conditions, and thus wouldn't exist if o didn't exist. For example:

> In uttering 'I'm sleepy' Sid meant the Sid-dependent proposition *that **he** was sleepy*, and therefore referred$_\Pi$ to himself in uttering that sentence.
> In uttering 'She divorced him', Jane meant the Angelina- and Brad-dependent proposition *that **Angelina** divorced **Brad***, and therefore referred$_\Pi$ both to Angelina and to Brad in uttering that sentence.

Higher-order speaker reference is exemplified by Sally's reference to Henry when she uttered 'That woman talking to Henry is French', where Odile is the woman to whom Sally referred and the proposition she meant was *that she, Odile, was French*. Here Odile was the only thing to which Sally referred$_\Pi$ in producing her utterance. Yet even though Sally didn't refer$_\Pi$ to Henry in uttering 'That woman talking to Henry is French', she nevertheless referred to him in uttering that sentence: she referred to Henry by virtue of her referring$_\Pi$ to Odile *qua woman standing next to Henry*. In other words, Sally didn't refer to Henry in order to mean something about him, but rather in order to identify the woman about whom she meant something. The reference to Henry is an instance of second-order speaker-reference: Sally referred to Henry in order to refer$_\Pi$ to Odile. Now consider Jack's utterance of 'The boy with the dog who's growling at Hilda is lost', where Billy is the boy to whom Jack referred$_\Pi$ and the proposition he meant was *that he, Billy, was lost*. Then Jack made a primary reference (i.e. a reference$_\Pi$) to *the boy with the dog who's growling at Hilda*, a secondary reference to *the dog who's growling at Hilda*, and, by virtue of referring to the dog *qua dog growling at Hilda*, a tertiary reference to Hilda. In (1981) I offered a recursive definition of speaker-reference devised by Brian Loar and myself which accommodated all cases of speaker-reference, but I will now henceforth ignore higher-order speaker-reference, drop the subscript 'π', and proceed as though speaker-reference were speaker-reference$_\Pi$.

This brings us to the important distinction between referring-*in* and referring-*with*.[2] Consider these two acts of speaker-meaning.

> In singing "Let the Good Times Roll," Gretel meant *that Hansel, her husband, was out of town*.

---

[2] In (1981) and elsewhere I spoke e.g. of *Jane's referring to John **by** (her utterance of) 'he'*. I thank Neale for suggesting the improvement of replacing 'by' with 'with'.

> In uttering 'He's out of town', Gretel meant *that Hansel, her husband, was out town.*

In both examples Gretel refers to Hansel *in* producing her utterance. In the first example she refers to him in singing 'Let the good times roll!', and in the second she refers to him in uttering 'He's out of town'. There is, however, the following important difference between the two utterances. In the first there is no part of Gretel's utterance *with which* she refers to Hansel, but there is in her second utterance: in that utterance she refers to him with her uttered token of 'he'. The following is a slightly revised version of the first approximation to defining this notion to which Neale appeals in "Silent Reference":

> In uttering x S referred to o with S's i^{th} token of e in x, $e_{\tau ix}$, iff for some audience A and feature φ, in uttering x S intended S and A to mutually know that $e_{\tau ix}$ has φ and, at least partly on the basis of this, that S referred to o in uttering x.

In other words, S referred to o with the token δ_τ of the singular term δ contained in her uttered token of the sentence σ when, roughly speaking, S intended δ_τ to have a certain feature φ and for the fact that δ_τ had φ to do most of the work in making it mutual knowledge between S and her audience that she referred to o in uttering σ. Of course, in (what is supposed to be) the normal case the inference-base feature φ will have as its most essential ingredient the fact that δ has the meaning it has in the shared language of S and her audience. For the Gricean the *raison d'être* of spoken languages is as instruments of interpersonal communication. If I want to tell my child that whales are mammals I can hardly do better than to utter 'Whales are mammals', and that, according to the Gricean, is because of the way the meaning of that sentence makes the fact that I uttered the sentence such extremely good evidence of my communicative intentions. That it is such good evidence is due to the meaning-determining conventional practices that prevail among speakers of the language to which the uttered expressions belong. I'll have a little more to say about this later.

Turning to expression-reference, the Gricean will find it useful to distinguish between a thing's being the *referent* and its being the *semantic* referent of a token of a singular term, where being the semantic referent entails being the referent, but not vice versa. The Gricean can define the first notion thus:

> x is the referent of a token δ_τ of a singular term δ iff the person who produced δ_τ referred to x with it.

For example, if a wedding guest points to a man in a tuxedo and says 'I asked that waiter for more champagne', then the indicated man is the referent of the uttered token of 'that waiter' whether or not he's a waiter or merely the groom's best man whom the wedding guest mistakes for a waiter. And he can define the second notion thus:

x is the semantic referent of a token δ_τ of a singular term δ iff (i) the person who uttered δ_τ referred to x with it and (ii) referring to x with δ_τ comports with the meaning of δ.

Thus, if the man to whom the wedding guest referred with his utterance of 'that waiter' wasn't a waiter, then, while the indicated man was the referent of the speaker's uttered token of 'that waiter', he wasn't the semantic referent of it, since, not being a waiter, the utterance of 'that waiter' didn't comport with the meaning of 'that waiter', since that meaning constrains one who utters the expression in conformity with the meaning-determining conventional practices of the language to be referring to a waiter. I'll have a little more to say below on how the Gricean conceives expression-meanings, and of how those meanings are supposed to constrain what speakers can mean in uttering sentences containing expressions with those meanings, but enough has been said for now for us to see how vague speaker-meaning makes trouble for the Gricean account of speaker-meaning and the semantic notions the Gricean aims to define in terms of it.

2. *Vague Speaker-Meaning*

Here is an unexceptional example of vague speaker-meaning. Tom is reading in the park when a woman appears, calling 'Billy, where are you? We have to leave now'. Intending to tell her something she might find helpful, Tom says to the woman 'A boy was here a little while ago'. We would certainly regard Tom as having told the woman something, and therefore as having meant and said something in producing his utterance. If the woman to whom he spoke didn't catch his words and asked him what he said, Tom wouldn't hesitate to say 'I said that a boy was here a little while ago', and we, knowing what we do, would accept Tom's report of what he said as true. We would unhesitatingly take Tom's utterance to be an act of speaker-meaning. Tom's utterance was also vague. His utterance was vague because even if it was definitely true or definitely false, it might have been indefinite whether or not it was true/false; that is to say, it might have been borderline true/false. It's three-ways overdetermined that Tom's utterance was vague, for its contained utterances of 'boy', 'here', and 'a little while ago' were vague, and the vagueness of any one of those utterances sufficed to make Tom's utterance of 'A boy was here a little while ago' vague. And each of those utterances was vague because, even if the application to which Tom put it was definitely correct or definitely incorrect, it might have been neither. I intend Tom's utterance to be an exemplar of vague speaker-meaning.

As already noted, the Gricean conception of speaker-meaning, as well as the dominant conception of it, is that of a relation between speakers and the propositions they mean. Let's assume *pro tem* that this conception of speaker-meaning is correct and ask: *What proposition did Tom*

mean in uttering 'A boy was here a little while ago'? Since propositions are mind- and language-independent entities, we can't answer this question without taking a stand on the issue of ontic vagueness, or vagueness in the world. That issue is a contest between a view I'll call *no-vagueness-in-the-world* and one I'll call *vagueness-in-the-world*. What exactly is at issue in this contest is itself in need of precisification, but to a good-enough approximation we may say that no-vagueness-in-the-world holds that nothing outside of language and thought can be vague in its own right (i.e. independently of the vagueness of language and thought), while vagueness-in-the-world holds that properties and things outside of language and thought may be vague in their own right. As a terminological convenience I'll say that things that aren't vague in their own right are *metaphysically precise*, and that things that are vague in their own right are *metaphysically vague*. A *proposition* is metaphysically precise provided it's necessarily the case that there is a fact of the matter as to what truth-value, if any, it has. If bivalence holds for propositions, then a proposition is metaphysically precise provided it's necessarily the case that it's a fact that it's true or else a fact that it's false. If, as Frege held, presupposition failure can render a proposition neither true nor false, then a proposition is metaphysically precise provided it's necessarily the case that it's a fact that it's true, a fact that it's false, or a fact that it's neither true nor false. And if there are three or more truth-values, then a proposition is metaphysically precise provided it's necessarily the case that there is a fact of the matter as to which truth-value it has. For simplicity of exposition I will assume that metaphysically-precise propositions are bivalent, either true or else false. As a matter of terminology I'll say that if a proposition is metaphysically-precise, then it can't be *metaphysically indefinite* what it's truth-value is, but that will be compatible with its being indefinite what it's truth-value is in an epistemic sense of 'indefinite', which is to say, the sense that it would have if the epistemic theory of vagueness were correct. A proposition is metaphysically-vague provided it can be metaphysically indefinite what it's truth-value is; in other words, if the proposition can be borderline true/false in its own right, independently of the vagueness of words and concepts. A property is metaphysically precise provided it's necessarily the case that everything is such that it either has the property or else doesn't have it, and a thing that is neither a property nor a proposition—e.g. an apple, a dog, a geographical area, or a period of time—is precise provided it's necessarily the case that it has metaphysically-precise conditions of individuation (so if it's a geographical area it can't be metaphysically indefinite what its boundaries are, and if it's a period of time it can't be metaphysically indefinite when it began or when it ended, or how many seconds or yoctoseconds[3] have passed between the instant of time it began and the instant of time it ended). Anything that isn't metaphysically precise is metaphysically vague.

---

[3] One yoctosecond = one trillionth of a trillionth of a second.

There is one possible answer to the question of what proposition Tom meant in uttering 'A boy was here a little while ago' with which we needn't be concerned—namely, that for some metaphysically-precise proposition p, Tom definitely meant p in uttering 'A boy was here a little while ago'. We needn't be concerned with this possible answer because if it were true Tom's utterance of 'A boy was here a little while ago' wouldn't have been vague. There remain four possible answers that are compatible with the vagueness of Tom's utterance. The first two are on the side of no-vagueness-in-the-world, the second two on the side of vagueness-in-the-world. That needs to be qualified. There are actually infinitely many possible answers—but we need to be concerned with only four of them. It's so-called *higher-order vagueness* that is responsible for the profligate proliferation. Consider Harold, whom we take to be a borderline case of a bald man, which is to say, nearly enough, that we take it to be indefinite whether or not he's bald. Now, the notion of a borderline case is itself vague; so there's the apparent possibility that he is a borderline case of a borderline case of a bald man, or a borderline case of a borderline case of a borderline case of a bald man, or …. The relevance of this to my argument may be illustrated in the following way. One of the possible answers I consider to the question "What proposition did Tom mean in uttering 'A boy was here a little while ago'?" is the following:

> There is no metaphysically-vague proposition that Tom definitely meant in uttering 'A boy was here a little while ago', but there are myriad metaphysically-vague propositions each such that it's indefinite whether or not Tom meant it in uttering that sentence.

If we now take higher-order vagueness into account, then other possible answers are:

> There is no metaphysically-vague proposition that Tom definitely meant in uttering 'A boy was here a little while ago', but there are myriad metaphysically-vague propositions each such that it's indefinite whether or not it's indefinite whether or not Tom meant it in uttering 'A boy was here a little while ago'.

> There is no metaphysically-vague proposition that Tom definitely meant in uttering 'A boy was here a little while ago', but there are myriad metaphysically-vague propositions each such that it's indefinite whether or not it's indefinite whether or not it's indefinite whether or not Tom meant it in uttering 'A boy was here a little while ago'.

> And so on without end.

That infinite sequence isn't the only way indefiniteness can proliferate. Another apparent type of possibility would be the series:

> There is a proposition p that Tom definitely meant in uttering 'A boy was here a little while ago' but it's indefinite whether or not p is metaphysically precise.

There is a proposition p that Tom definitely meant in uttering 'A boy was here a little while ago' but it's indefinite whether or not it's indefinite whether or not p is metaphysically precise.

There is a proposition p that Tom definitely meant in uttering 'A boy was here a little while ago' but it's indefinite whether or not it's indefinite whether or not whether or not it's indefinite whether or not p is metaphysically precise.

And so on without end.

And one can easily see that other kinds of infinite proliferations are possible. I propose, however, to ignore higher-order vagueness altogether and pretend that it's only first-order vagueness with which we have to contend, and I will assume it's always metaphysically definite whether a proposition (or anything else) is metaphysically precise or metaphysically vague; for I trust it will be clear how the objections I will raise to the four possible answers I do consider also apply to the infinitely many possible answers I don't consider.

The four possible answers to be considered are as follows.

(A)   For some metaphysically-precise proposition p, Tom meant p in uttering 'A boy was here a little while ago', but it's indefinite whether or not he meant p in uttering that sentence.

(B)   There is no metaphysically-precise proposition that Tom definitely meant in uttering 'A boy was here a little while ago', but there are myriad metaphysically-precise propositions each such that it's indefinite whether or not he meant it in uttering that sentence.

(C)   For some metaphysically-vague proposition p, Tom meant p in uttering 'A boy was here a little while ago'.

(D)   There is no metaphysically-vague proposition that Tom definitely meant in uttering 'A boy was here a little while ago', but there are myriad metaphysically-vague propositions each such that it's indefinite whether or not he meant it in uttering that sentence.

The big question now is whether any of (A)-(D) can survive scrutiny.

Re (A) [For some metaphysically-precise proposition p, Tom meant p in uttering 'A boy was here a little while ago', but it's indefinite whether or not he meant p in uttering that sentence]. If it's indefinite whether or not such-and-such is the case, then it's impossible for anyone to know whether or not such-and-such is the case. But how can there be any proposition that Tom meant in uttering 'A boy was here a little while ago' if it's impossible for him or anyone else to know that he meant it? Yet (A) would be true if the epistemic theory of vagueness were true, for that theory expounds the thesis that "the proposition a vague sentence expresses in a borderline case is true or false, and we cannot know which" (Williamson 1997: 921). There are issues about how we should understand the ignorance about borderline cases epistemicism

requires, and I'll have a little to say about that presently, but even without resolving those issues we can know that (A) is false if we can know that there couldn't have been a metaphysically-precise proposition that Tom meant in uttering 'A boy was here a little while ago', and I believe we can know that in the following way.

The vagueness of Tom's utterance, we have noticed, was three-ways overdetermined: by the vagueness of his utterance of 'boy', the vagueness of his utterance of 'here', and the vagueness of his utterance of 'a little while ago'. Consequently, if Tom meant a metaphysically-precise proposition in uttering 'A boy was here a little while ago', then it must also have been the case that:

(1)   for some metaphysically-precise property φ, Tom expressed φ with the token of 'boy' he uttered (if we pretend that 'male' and 'human being' express metaphysically-precise properties and that there is a metaphysically-precise moment at which a person comes into existence, then φ might be the property of being a male human being whose age in milliseconds ≤ 531,066,240,000);[4]

(2)   for some metaphysically-precise area α—i.e. area that has metaphysically-precise boundaries, and thus comprises a metaphysically-precise number of square millimeters, and a metaphysically-precise location relative to Tom's location— Tom referred to α with 'here';

(3)   for some metaphysically-precise period of time π, Tom referred to π with 'a little while ago', where in order for that to be so there must be instants of time of 0 duration t, t′, t′′, and real numbers n, n′, n′′, such that (i) t is the instant π began, (ii) t′ is the instant π ended, (iii) t′′ is the instant of time a "little while" before which Tom is saying the boy was in α, the metaphysically-precise area to which Tom referred with 'here', (iv) n is the precise number of milliseconds between t and t′, (v) n′ is the precise number of milliseconds between t′′, the instant of time from which all measurements of time relevant to the reference of 'a little while ago' emanate, and t, the instant of time such that if the boy's appearance in α was so much as one yoctosecond before t, than it was too long before t′′ to count as "a little while ago," and (vi) n′′ is the precise number of milliseconds between t′′ and t′, the end of π and thus the instant of time such that if the boy's appearance in α was so much as one yoctosecond after t′, then it was too soon before t′′ to count as a "little while *ago*."

We can show that Tom didn't mean any metaphysically-precise proposi-

---

[4] As an expository convenience, instead of saying e.g. *Tom expressed φ with the token of 'boy' he uttered* I'll say *Tom expressed φ with 'boy'*, where that will be shorthand for the longer way of speaking. Likewise *Tom referred to α with 'here'*, for example, will be shorthand for *Tom referred to α with the token of 'here' he uttered*.

tion in uttering 'A boy was here a little while ago' if we can show that any one of (1)–(3) is false. I think we can show that each of (1)–(3) is false, but I'll begin with (2) since it seems to be the simplest of the three requirements. There are at least the following reasons why Tom couldn't have referred to a metaphysically-precise area, whether or not a Gricean account of speaker-reference is correct.

(i)   On *any* plausible account of speaker-reference, acts of speaker-reference are *intentional* acts in that, if for some S, o and x, S referred to o with x, then S uttered x with those o-directed intentions that are constitutive, or at least partly constitutive, of her referring to o with x, and if S has the concept of speaker-reference, then, in the normal case, she intended to refer to o with x. For the Gricean this is so in spades. Tom's utterance was a normal case, but it ought to be obvious that there was no metaphysically-precise area to which Tom intended to refer when he produced it, nor would he have thought there was any need to refer to such an area. We should expect a fuller description of the imagined scenario to include the fact that when he uttered 'A boy was here a little while ago', Tom was confident that he was saying something true because he was confident that a human male child no older than six was within four meters of him no more than ten minutes before he spoke. Tom gave no thought to which of the uncountably many metaphysically-precise areas containing the boy was the one he wanted to make a statement about, because there was no metaphysically-precise area about which he wanted to make a statement. Careful and considerate speakers try not to use a vague term unless they are confident that their use of it would be recognized as correct, but they have no reason to consider what would have to have been the case for their use of the term to have been borderline correct, or just barely correct or incorrect. Tom, if he understands the question and has his wits about him, ought to be rendered speechless by the question "Exactly which area did you intend to refer when you uttered 'here'?" Nor did Tom think there was any need to refer to any particular area that contained the boy, for he would think he succeeded in giving the woman the information he intended to give her if, as a result of his utterance, she believed that a boy was at a location within the vicinity of Tom which made his utterance true, and that didn't require her to think that any particular area in which the boy was contained was the area to which Tom referred in producing his utterance.

(ii)   It's both a cornerstone of the Gricean account of speaker-reference and obviously true that we refer to things in order to make known to our hearers what we are talking about. In a normal case, such as Tom's, a speaker can't refer to a thing if she knows that her hearer wouldn't be able to know to what she was refer-

ring. Tom would know that even if there were a metaphysically-precise location to which he wanted to refer, his hearer would have no way of knowing which of the uncountably many eligible metaphysically-precise areas was the one to which he was referring. Given that, he couldn't have intended to refer to any metaphysically-precise area.

(iii)   There is a deeper explanation of why Tom couldn't have referred to any metaphysically-precise area. The statement

There is a metaphysically-precise area α such that Tom intended to refer to α with 'here', if true, ascribes to Tom an intention that is *de re* with respect to an unspecified area α, and just as one can have a belief that is *de re* with respect to a thing under one way of thinking of it but not under another, so one can have an intention that is *de re* with respect to a thing under one way of thinking of it but not under another. Let $α^*$ be any metaphysically-precise area in the vicinity of Tom . What way of thinking of $α^*$ might Tom have under which it would be possible for him to intend to refer to $α^*$ with 'here'? He has no perceptual way of thinking of $α^*$ that would do the job, and it's very difficult to see what knowledge by description of $α^*$ he might have that would enable him to intend to refer to $α^*$ under it.[5] It seems impossible to think of any kind of way of thinking of a metaphysically-precise area that would yield a way of thinking of $α^*$ under which Tom might have *any* intention or belief that was *de re* with respect to it. In short, it seems that there couldn't have been anything about any particular metaphysically-precise area that would explain what made it, rather than any of the uncountably other metaphysically-precise areas that differed only imperceptibly from it, the area to which Tom referred with 'here'.

I conclude that there was no metaphysically-precise area to which Tom referred in uttering 'A boy was here a little while ago', and that entails that there was no metaphysically-precise proposition that Tom meant in uttering 'A boy was here a little while ago'. So (A) is false. The same sort of considerations used to show that Tom couldn't have referred to a metaphysically-precise area with 'here' can also be used to show that he couldn't have referred to a metaphysically-precise period of time with 'a little while ago'. In fact, given the complexity of what would have to be the case in order for Tom to have referred to a metaphysically-precise period of time (see above p. 302), it should be more intuitively obvious that he couldn't have referred to a metaphys-

---

[5] If factors that didn't require Tom to have any intention that was *de re* with respect to $α^*$ secured it as the referent of the token of 'here' Tom uttered, then perhaps Tom could have intended to refer to $α^*$ under the description *the area to which the token of 'here' I uttered refers*. Yes, *but* (1) it's impossible to see what factors could determine that reference if they didn't include a *de re* intention of Tom's about $α^*$ and (2) it's just as difficult to see what feature one of uncountably many indiscriminable metaphysically-precise areas could make it alone the referent of the token of 'here' as it is to see what feature could make it alone the one to which Tom referred with that token of 'here'.

ically-precise period of time than it is that he couldn't have referred to a metaphysically-precise area. So that's another way to show that there couldn't have been a metaphysically-precise proposition which Tom meant in uttering 'A boy was here a little while ago', and therefore another way to show that (A) is false.

Can the same sort of considerations used to show that Tom couldn't have referred either to a metaphysically-precise area with 'here' or to a metaphysically-precise period of time with 'a little while ago' also be used to show that Tom couldn't have expressed a metaphysically-precise property with 'boy'? I believe so, but it's a little trickier to show this. We can see why it's trickier in the following way.

Some theorists would say that the vagueness of 'boy' is no barrier to the truth or knowability of claims such as

(i)   The meaning of 'boy' = the property of being a boy

or

(ii)   'Boy' is true of a person iff he is a boy

For, they would say, the vagueness of the right-hand side occurrence of 'boy' in (i) and (ii) is simply the vagueness of the word—viz. 'boy'—referred to on the left-hand side, so that the vagueness of the one balances out the vagueness of the other. Suppose that is right, that the epistemic theory of vagueness is true, and that the property of being a boy is metaphysically-precise, so that, in addition to (i) and (ii) being true, a statement like the following is also true:

(iii)   The property of being a boy = the property of being a male human being whose age in milliseconds seconds ≤ 531,066,240,000.

Let's pretend that the statement like (iii) that is true is (iii) itself. Then (i) and (ii) would be equivalent, respectively, to

(iv)   The meaning of 'boy' = the property of being a male human being whose age in milliseconds ≤ 531,066,240,000

and

(v)   'Boy' is true of a person iff he is a male human being whose age in milliseconds ≤ 531,066,240,000.

Now, according to the epistemic theorist, in order for (iii), (iv), and (v) to be compatible with the vagueness of 'boy' it would have to be impossible for anyone to know any one of them, for if they could know any one of them then it couldn't ever be indefinite whether a human male was a boy, since it would always be in principle possible to determine whether or not a male human was a boy by computing the number of milliseconds that have passed since he was born. But the unknowability of the truths (iii)–(v) wouldn't render the truths (i) and (ii) unknowable, for just as one can know that George Eliot wrote *Middlemarch* but not know that Mary Ann Evans wrote *Middlemarch*, even though George Eliot was Mary Ann Evans, so, it might be said, one can know

(i) and (ii) even if it was impossible for one to know (iii), and therefore impossible for one to know (iv) and (v). The upshot of this would be that, if the epistemic theory is true, then the difficulty in a speaker's meaning a metaphysically-precise proposition in producing a vague utterance arises only for vague utterances that include vague utterances of expressions such as 'here' and 'a little while ago', but not for vague utterances of terms like 'boy' whose meanings (we might suppose) are metaphysically-precise things or properties. That line of thought, then, is why showing that Tom couldn't have expressed a metaphysically-precise property with 'boy' is trickier than showing that he couldn't have referred to a metaphysically-precise area with 'here' or to a metaphysically-precise period of time with 'a little while ago'.

In fact, however, the line of thought is specious; Tom's utterance of 'boy' really is in the same boat as his utterances of 'here' and 'a little while ago'. If 'boy' did mean the property of being a boy, then it would be easy for Tom to intend to express the property of being a boy with the token of 'boy' he uttered. But 'boy' can't mean the property of being a boy, and what prevents there being a metaphysically-precise property that Tom expressed with 'boy' is on all fours with what prevents him him from referring to a metaphysically-precise area with 'here' or to a metaphysically-precise area with 'a little while ago'. This is due to a feature of every vague expression which precludes any kind of thing or property from being the meaning of any vague expression, a feature which shows that *none* of (i)–(v) is compatible with the vagueness of 'boy', even if the epistemic theory is true. That feature is one I call *penumbral shift*.[6] Penumbral shift doesn't per se show that speakers don't mean metaphysically-precise propositions in acts of vague speaker-meaning, but it does show that Tom's utterance of 'boy' makes it no less difficult for him to have meant a metaphysically-precise proposition in uttering 'A boy was here a little while ago' than does his utterance of 'here' or of 'a little while ago'.

Penumbral shift is a feature of every vague expression, but to economize on words I will explain it only with respect to vague predicates like 'boy', 'violinist', 'mathematician', 'communist', and 'triangular' which, unlike such vague predicates as 'local', 'ready', 'tall' or 'intelligent', are apt to be regarded as having constant characters if vagueness is being ignored. I'll refer to such vague predicates as predicates*. Now, every token of a predicate* has a *penumbral profile*, and two tokens of a predicate* have the same penumbral profile just in case if either token is true/false of a thing, then likewise for the other; if either token is such that it's indefinite whether or not it's true/false of a thing, then likewise for the other; if either token is such that it's indefinite whether or not it's indefinite whether or not it's true/false of a thing, then likewise for the other; and so on. Then we may say that penumbral shift (when restricted to predicates*) is the fact that *the penumbral profiles*

---

[6] Schiffer (2010) and (2016).

of tokens of a predicate* may shift somewhat from one token of the predi-cate* to another; that is to say, two tokens of a predicate* may have somewhat different penumbral profiles. The "somewhat" qualification is important. For example, if Clyde is a man whose scalp is as hairless as a billiard ball and on whose scalp no hair can grow, then every token of 'bald man' must be true of Clyde, but if Clyde is blessed with a head of hair like the one Tom Cruise appears to have, then every token of 'bald man' must be false of him. At the same time, penumbral shift makes it possible for there to be three tokens of any predicate*, one of which is true of the thing to which it's applied, another of which is false of the thing to which it's applied, and still another of which is such that it's indefinite whether or not it's true of the thing to which it's applied. Here are three examples:

— At a party George is asked whether Henrietta came to the party with anyone, and he replies, gesturing at a certain man, 'She came with that bald guy'. That utterance would most likely be accept-ed in the context in which it occurred as true, even if it somehow transpires that the man in question shaves his scalp but wouldn't be said to be bald if he let his hair grow out. In another conversa-tion, however, in which the discussion is about hereditary baldness, someone might correct a remark about the same man by saying, 'No; he's not bald; he just shaves his scalp', and that utterance, in that context, would very likely count as true. In still another context the question is raised whether a man who shaved his scalp would be bald if no one would take him to be bald if he stopped shaving his scalp and let his hair grow out, and in that context it might be true to say, 'That's undetermined by the use of "bald" in everyday speech; such a man would be neither definitely bald nor definitely not bald'.

— In a community in which people typically marry before the age of twenty, an utterance of 'He's a bachelor' may count as true when said of an unmarried eighteen-year-old male, whereas in a conver-sation among New Yorkers, where for both men and women the av-erage age for a first marriage is between thirty and thirty-five, an utterance of 'He's a bachelor' would most likely not count as defi-nitely true when said of an unmarried eighteen-year-old male, and may even count as false.

— An utterance of 'Mary is getting married to a boy from Boston' would count as false if the male whom she is about to marry is fifty-two years old, but is apt not to count as false if he is twenty-seven years old. At the same time, if every one of the seven male profes-sors in one's department is over forty except Henry, who is twenty-seven, an utterance of 'Six men and one boy are professors in my department' would count as a misuse of 'boy'. The boys' clothing section in a department store isn't where a normal-size eighteen-year-old American male would go to buy clothes, but if Jack is an

eighteen-year-old high school student who is the star of his school's boys basketball team, an utterance of 'Jack is the tallest boy on the team' would count as true if he is the tallest person on the team.

Now we can see why, owing to the vagueness of 'boy', penumbral shift makes it impossible for any of (i)–(v) to be true. It precludes

(i)   The meaning of (the predicate-type) 'boy' = the property of being a boy

from being true because:

1)   Necessarily, the just-displayed *token* of (i) is true only if, for some property φ, φ is the referent of the token of 'the property of being a boy' in that token of (i), and the predicate-type 'boy' means φ.
2)   Necessarily, if a property φ is the meaning of a predicate-type F, then F has a constant character and φ is the content of every token of F.
3)   Necessarily, if a property φ is the content of every token of a predicate F, then there aren't simultaneously two tokens of F one of which is true of a thing, the other of which isn't true of that thing.
4)   If a predicate F is subject to penumbral shift, then there can be two simultaneous tokens of F one of which is true of a thing, the other of which isn't true of that thing.
5)   'Boy', being vague, is subject to penumbral shift.
6)   So, no token of (i) can be true.

A similar argument shows that penumbral shift also precludes any token of

(ii)   'Boy' is true of a person iff he is a boy

from being true. From here it should be easy to see that each of

(iii)   The property of being a boy = the property of being a male human being whose age in milliseconds $\leq 531{,}066{,}240{,}000$.

(iv)   The meaning of 'boy' = the property of being a male human being whose age in milliseconds $\leq 531{,}066{,}240{,}000$.

(v)   'Boy' is true of a person iff he is a male human being whose age in milliseconds $\leq 531{,}066{,}240{,}000$.

is also incompatible with the fact that 'boy', being vague, is subject to penumbral shift. None of this shows that penumbral shift is per se incompatible with the epistemic theory of vagueness. It only shows that penumbral shift is incompatible with a version of the epistemic theory which holds that some property is the meaning of 'boy'. The upshot of all this as regards the hypothesis that Tom meant a metaphysically-precise proposition is that, whether or not the epistemic theory is true, the problem that his utterance of 'here' or 'a little while ago' makes for the proposal that Tom meant a metaphysically-precise proposition in

uttering 'A boy was here a little while ago' is equally made by his utterance of 'boy'. For if Tom meant a metaphysically-proposition, then there was a metaphysically-precise property that Tom expressed with 'boy', and now that we see that that property can't be the meaning of 'boy', it will be just as difficult to see how just one of the nearly identical metaphysically-precise properties in contention—e.g. *the property of being a human male whose age in milliseconds ≤ 531,066,240*, 000, as opposed, say, to *the property of being a human male whose age in milliseconds ≤ 531,066,239,000.07* or *the property of being a human male whose age in milliseconds ≤ 531,066,240,000.8*— could be determined to be the metaphysically-precise property expressed by Tom's uttered token of 'boy' as it was to see how just one of the metaphysically-precise areas in contention could be determined to be the metaphysically-precise area to which he referred with the token of 'here' he uttered.

I conclude that we know that there was no metaphysically-precise proposition that Tom meant in uttering 'A boy was here a little while ago', and therefore know that (A) is false.

Re (B) [There is no metaphysically-precise proposition that Tom definitely meant in uttering 'A boy was here a little while ago', but there are myriad metaphysically-precise propositions each such that it's indefinite whether or not he meant it in uttering that sentence]. We know from the discussion of (A) that there was no metaphysically-precise proposition that Tom meant in uttering 'A boy was here a little while ago', and if we know that, then it's definitely the case. Therefore, it's not the case that there are myriad metaphysically-precise propositions each such that it's indefinite whether or not Tom meant it in uttering 'A boy was here a little while ago', for the considerations adduced to show that (A) is false also show that every metaphysically-precise proposition is such that Tom definitely did not mean it. Therefore, (B), as well as (A), is false.

Re (C) [For some metaphysically-vague proposition p, Tom meant p in uttering 'A boy was here a little while ago']. This answer presupposes vagueness-in-the-world; according to it Tom's utterance was vague because he meant a metaphysically-vague proposition in producing it. Many philosophers will doubt (C) because they doubt whether the notion of metaphysical vagueness can be made coherent. Here is how David Lewis expressed his own frustration with the idea that there are metaphysically vague things:

> I doubt that I have any correct conception of a vague object. How, for instance, shall I think of an object that is vague in its spatial extent? The closest I can come is to superimpose three pictures. There is the multiplicity picture, in which the vague object gives way to its many precisifications, and the vagueness of the object gives way to differences between precisifications. There is the ignorance picture, in which the object has some definite but secret extent. And there is the fadeaway picture, in which the presence of the object admits of degree, in much the way that the presence of a spot of illumination admits of degree, and the degree diminishes as a function

of the distance from the region where the object is most intensely present. None of the three pictures is right. Each one in its own way replaces the alleged vagueness of the object by precision. But if I cannot think of a vague object except by juggling these mistaken pictures, I have no correct conception. (Lewis 1999a: 170)

That objection to vagueness-in-the-world is certainly discussable, but in order to give the propositional account of speaker-meaning its best run for the money, I will for present purposes assume that vagueness-in-the-world is at least coherent. My question, then, concerns the plausibility of (C) on the assumption that there are metaphysically-vague properties, areas, periods of time, and propositions. The answer to my question, I submit, is that (C) isn't plausible even on that assumption: it's shown to be false by exactly the same sort of considerations that showed (A) to be false.

The hypothesis that Tom meant a metaphysically-*precise* proposition in uttering 'A boy was here a little while ago' entails that he expressed a metaphysically-precise property with 'boy', referred to a metaphysically-precise area with 'here', *and* referred to a metaphysically-precise period of time with 'a little while ago'. The hypothesis that Tom meant a metaphysically-*vague* proposition in producing his utterance entails only that he expressed a metaphysically-vague property with 'boy', referred to a metaphysically-vague area with 'here', *or* referred to a metaphysically-vague period of time with 'a little while ago'. For example, it's compatible with Tom's having a meant a metaphysically-vague proposition that he expressed a metaphysically-vague property with 'boy' but referred to a metaphysically-precise area with 'here' and to a metaphysically-precise period of time with 'a little while ago'. At the same time, as we have seen, we can't account for the vagueness of Tom's utterances of 'here' and 'a little while ago' if he referred to metaphysically-precise things with those expressions, so in fact the only way he could have meant a metaphysically-vague proposition that was compatible with the vagueness of his utterances of 'boy', 'here', and 'a little while ago' is if he expressed a metaphysically-vague property with 'boy', referred to a metaphysically-vague area with 'here' and to a metaphysically-vague period of time with 'a little while ago'. The upshot as regards (C) is that we can see that it's false by seeing that Tom couldn't have referred to any metaphysically-vague area for the same reason, mutatis mutandis, that he couldn't have referred to any metaphysically-precise area. The fundamental reason Tom couldn't refer to any metaphysically-precise area is that he had no way of thinking about any such area under which he might intend to refer to it, and this because each such area was for him indistinguishable from the uncountably many precise areas that differed from it only in some imperceptible way. The same is also true of the uncountably many metaphysically-vague areas in Tom's vicinity (if there are such things). For example, there will be two such areas α_1 and α_2 such that if any location is definitely in α_1, then it's also definitely in α_2, and vice versa,

the only difference between the areas being that there are locations such that while it's indefinite whether or not they are in α_1, it's merely indefinite whether or not it's indefinite whether or not they are in α_2, and that's not a difference that would enable Tom to intend to refer to either area, for he would still have neither a perceptual nor descriptive way of thinking of either area under which he could have any *de re* propositional attitudes about it. If there are metaphysically-vague things or properties, they are every bit as finely individuated as any metaphysically-precise thing or property.

Re (D) [There is no metaphysically-vague proposition that Tom definitely meant in uttering 'A boy was here a little while ago', but there are myriad metaphysically-vague propositions each such that it's indefinite whether or not he meant it in uttering that sentence]. (D) stands to (C) as (B) stands to (A). We know from the discussion of (C) that there was no metaphysically-vague proposition that Tom meant in uttering 'A boy was here a little while ago', and if we know that, then it's definitely the case. Therefore, it's not the case that there are myriad metaphysically-vague propositions each such that it's indefinite whether or not Tom meant it in uttering 'A boy was here a little while ago', for the considerations adduced to show that (C) is false also show that, if there are metaphysically-vague propositions, then every metaphysically-vague proposition is such that Tom definitely did *not* mean it. Therefore, (D), as well as (C), is false.

3. *The Cost of Vague Speaker-Meaning for Gricean Semantics*

So, I submit, none of (A)–(D) survives scrutiny, and from this we may conclude that there was no proposition that Tom meant in uttering 'A boy was here a little while ago'. Tom's utterance is an arbitrary example of vague speaker-meaning, and since virtually every utterance is vague, virtually every utterance is an instance of vague speaker-meaning. If there is no proposition that Tom meant in producing his utterance, then speakers virtually never mean propositions when they speak. Let's explore the consequences of this for Gricean semantics under three headings: *speaker-meaning*, *propositional attitudes*, and *meta-semantics & semantics*.

Speaker-meaning (and propositional speech acts generally)

The Gricean, as well as the dominant, conception of speaker-meaning is as a relation between speakers and the propositions they mean. This conception collapses in the face of vague speaker-meaning, if what I've argued in the preceding section is correct. This of course doesn't mean that an utterance such as

In uttering 'Phil and Barbara have three kids', Sid meant that Phil

and Barbara have three young goats

can't be true, but it does mean that if it's true then its 'that'-clause doesn't refer to any proposition. This raises two questions: First, might the Gricean program, or at least some version of it, make do with something other than propositions to be the things we mean, say, and tell people? Second, should any other erst-while propositionalist seek something other than propositions to be the things we mean? I believe the answer to both questions is *no*. If there are such things as the things we mean, then those things must be capable of having truth-values, and if those things aren't propositions, then they must either be mentalese sentences or, more likely, public language sentences or utterances (think of Donald Davidson's "paratactic" theory of saying-that). The mentalese version of sententialism has a host of problems that keep it from being a serious option for anyone,[7] and the public-language version is unacceptable to the Gricean because it requires a non-Gricean account of the semantic properties of expressions.[8] But there is another reason why there aren't things of any kind that can be the things we mean if, for the reasons I've given, propositions can't be the things we mean—viz. there would have been nothing to prevent there having been a proposition that Tom meant in uttering 'A boy was here a little while ago' if there had been a metaphysically-precise vague property, area and period of time to which he referred with, respectively, the tokens he uttered of 'boy', 'here', and 'a little while ago', but, as I pointed out in fn. 6, it's just as difficult to see, for example, what feature one of uncountably many indiscriminable metaphysically-precise or vague areas could make it alone the referent of the token of 'here' as it is to see what feature could make it alone the one to which Tom referred with that token of 'here'. I conclude that, while there are true reports like the one displayed just above, neither speaker-meaning nor any other other "propositional" speech act is a relation to anything. I'll presently say something about the challenge this conclusion poses.

Propositional attitudes

The dominant view of propositional attitudes, as well as the view of them Gricean semantics requires, is that they are…well, *propositional* attitudes. The dominant view of believing, for example, is that it's a relation between a believer and a proposition she believes. But if the considerations adduced to show that there is no proposition that Tom meant in producing his utterance really do show that, then they also show that vague propositional attitudes aren't relations to propositions. This is an important point. Other philosophers have made their own trouble for the view that communication involves a speaker's uttering words that encode the proposition she wants to communicate,

---

[7] See Schiffer (1987: Chapter 4 "Intentionality and the Language of Thought").

[8] See op. cit., Chapter 5 "Sententialist Theories of Belief."

and that her attempted communication is successful just in case her hearer successfully decodes the encoded proposition. But the philosophers who have argued against this view of communication have held that, while successful communication doesn't consist in a hearer's entertaining or believing the very same proposition that is the content of the belief the speaker expressed in producing her utterance—i.e. the belief that was the proximal cause of her utterance—it does consist in a certain similarity-relation's obtaining between the proposition the speaker believed and the one the hearer entertained or believed as a result of the speaker's utterance. That was the view to which Frege was giving voice when he wrote:

> In the case of an actual proper name such as 'Aristotle' opinions as to the sense may differ. It might, for instance, be taken to be the following: the pupil of Plato and teacher of Alexander the Great. Anybody who does this will attach another sense to the sentence 'Aristotle was born in Stagira' than will a man who takes as the sense of the name: the teacher of Alexander the Great who was born in Stagira. So long as the reference remains the same, such variations of sense may be tolerated, although they are to be avoided in the theoretical structure of a demonstrative science and ought not to occur in a perfect language.[9]

But if what I have been arguing is on the right track, not only was there no proposition that Tom meant in uttering 'A boy was here a little while ago', there were also no propositions that were the contents of the beliefs and intentions that lead him to utter that sentence. This should be clear, for if what I said about Tom's utterance of e.g. 'here' was correct, the reason he couldn't refer either to a metaphysically-precise or to a metaphysically-vague area with 'here' is that there was nothing about any area of either kind that could explain how Tom could have an intention or belief that was *de re* with respect it. At the same time, to redirect to propositional-attitude reports a point already made about speech-act reports, it remains true that what lead to his utterance was his knowing that a boy had been in his vicinity a little while before he spoke, and that he said what he did to the woman to whom he spoke because he wanted to share that knowledge with her. It's just that although the propositional-attitude report I just made in writing the preceding sentence is true, neither its 'that'-clause nor the occurrence of 'that knowledge' in it refers to a proposition.

[9] Frege (1892). For contemporary expressions of the view see for example McDowell (1984a), Heck (2002), Buchanan and Ostertag (2005), and Buchanan (2010). NYU Ph.D. student Martin Abreu defends a novel version of this line in his nearly completed doctoral dissertation.

Meta-semantics & semantics

How do the conclusions so far reached affect what the Gricean needs to say about the nature and determinants of the semantic properties of linguistic expressions? To answer this I'll begin by sketching what the Gricean needs to say about those things. My sketch will impose a certain degree of regimentation, and there are ways my sketch may be varied while remaining faithful to the Gricean program, but I believe that the sketch I'm about to give captures what must be regarded as essential to the program.

A language, as David Lewis liked to emphasize (Lewis 1969, 1975), is a certain kind of abstract object that may or may not be used by anyone—to wit, a finitely stable function that maps each of infinitely many sequences of sounds, or marks, or hand gestures, or whatever (the expressions of the language) to things that do the job that "meanings" are supposed to do. If a function L is a language and L(ε) = μ, then we may stipulate that ε is an "expression" of L and μ its "meaning" in L. What sort of thing a meaning must be is determined by the work a language must do in order for it to be the language of a given population. To say what that way is is to define a relation R such that, necessarily, if a language L bears R to a population P, then L is the language of P, where that is equivalent to saying that every expression of L means in P what it means in L, and where the notion of meaning-in-a-population is the use-dependent notion of meaning that philosophers have long struggled to understand. Let's call that relation, whatever it turns out to be, the *public-language relation*. The Gricean, we know, aims to define the public-language relation in terms of his defined notion of speaker-meaning, together with certain ancillary notions, such as that of a conventional practice, which are themselves definable in terms of ordinary beliefs and intentions and without recourse to any public-language semantic notions. So, if we know the Gricean's account of the public-language relation, then we will know exactly how a language must be used by a population in order for it to be the population's public language, the language members of the population use to communicate with one another, and in knowing that we will know what role something must play if it's to be the meaning an expression has for the members of that population. The following gives the gist of a Gricean account of the public-language relation:

> L is a language of P iff there prevails in P a system of conventional practices conformity to which requires one who produces an unembedded utterance of a sentence of L to mean thereby a proposition that (in a sense presently to be explained) "fits" the meaning of the sentence in L .[10]

---

[10] An unembedded utterance of a sentence is an utterance of it that doesn't occur as part of the utterance of another sentence.

This explains how for the Gricean uttering a sentence with a certain meaning can be extremely good evidence that in uttering the sentence the speaker meant a proposition that fits the sentence's meaning.

When we understand the "fitting" relation we will know what the Gricean takes expression-meanings to be. To understand that relation we must understand the meaning/content distinction. *Meanings* are properties of expression-*types*, *contents* of expression-*tokens*. The meaning of an expression-type is, in effect, a rule or formula that specifies the conditions that something must satisfy if it's to be the content of a token of the expression. Then we can say that the proposition a speaker means in uttering a sentence is the content of the token of the sentence the speaker uttered just in case the proposition satisfies the rule or formula that the sentence means. (Specifications of these meanings are stable in "character"-style as functions from tokens onto their contents.) Propositions are the contents of sentence tokens, and a primary role such a content plays is to be the proposition the speaker means when he utters an unembedded sentence token whose content is that proposition. The content of every expression token will be what it contributes to the content of the unembedded sentence token in which the expression token occurs. If a Gricean is ignoring vagueness, then to a rough approximation he might say that:

— The content of every token of 'boy' (that has a content) is the property of being a boy.

— The content of a token of 'here' is an area α such that the speaker who uttered the token is within α and she referred to α with the token of 'here' she uttered.[11]

— The content of a token of 'a little while ago' is a period of time π such that π occurred shortly before the token was uttered and the speaker who uttered the token referred to π with it.

— The content of a token of 'A boy was here a little while' is the proposition that something that had φ was in α at a time within π, where φ is the content of the uttered token of 'boy', α the content of the uttered token of 'here', and π the content of the uttered token of 'a little while ago'.

What remains of this Gricean semantic and meta-semantic picture when one tries to adjust it to accommodate vagueness? The answer, if what I've argued is correct, should be clear: *absolutely nothing.* A language won't be "a pairing of sound and meaning over an infinite domain" (Chomsky 1980: 82), for there won't be anything that can be the content of a token of a vague expression, and therefore nothing to be the meaning of a vague expression. At the same time, the Gricean doesn't stand alone, for, as I've already implied (see above p. 312), the

---

[11] This ignores the use of 'here' manifested when the FBI agent points to a certain spot on a map and says to the guy who's about to enter the witness protection program, 'Here is where you'll live for the foreseeable future'.

considerations that frustrate his attempt to assign denotations to the tokens of the vague expressions that occur in a token of the sentence 'A boy was here a little while ago' will frustrate the attempt of any other theorist to do the same.[12]

You might feel that I've painted myself into a corner I can't get out of. I've committed myself to the following two claims about Tom's utterance of 'A boy was here a little while ago'. First, that the utterance was true, and second, that it's truth-value wasn't a function of the denotations of its constituent expressions. But how can the utterance be true unless it has truth conditions, and how can it have truth conditions if they're not a function of the denotations the sentence's constituent expressions have in the utterance. Well, there is *a sense* in which the utterance has truth conditions and *a sense* in which those conditions, such as they are, are determined by denotation-like properties of the utterances of the sentence's constituent expressions. The utterance of the sentence has truth conditions in the sense that there are myriad states of affairs which, if realized, would make the utterance determinately true, and myriad states of affairs which, if realized, would make the utterance determinately false. An example of the first sort would be a state of affairs in which a five-year-old human male was within three meters of Tom eight minutes before he spoke, and an example of the second sort of state of affairs would be one in which a ninety-year-old woman was the only person to be within a kilometer of Tom in the two hours before he spoke. Moreover, that those states of affairs have that status is clearly due to there being conditions whose satisfaction by a use of 'boy', or 'here', or 'a little while ago' would make Tom's use of it determinately correct, likewise for conditions that would make the use of any one of those vague expressions determinately incorrect. What there can't be is a *set* of states of affairs such that the utterance was determinately true just in case some member of the set was realized, another set such that the utterance was determinately false just in case some member of that set was realized, or a set of states of affairs such that an utterance of the sentence would be such that it would be indeterminate whether or not it's true just in case some member that set was realized. And so on. Likewise, *mutatis mutandis*, for there not being for any vague expression a set of conditions such that an utterance of the expression would be determinately correct just in case some member of the set was satisfied. And so on. Nevertheless, my conclusions do seem to preclude a natural language from having a correct compositional meaning theory or a correct compositional truth theory, where a compositional meaning theory is taken to be a finitely statable theory that issues for each of the infinitely many expressions of the language a theorem that assigns to it its meaning in the language, and where a compositional truth theory for a language is taken

---

[12] I take this claim to be more or less obvious, but if it isn't (and even if it is) I can't take the space here to give the claim the elaboration it may deserve.

to be a finitely stable theory of the language that issues for each of the infinitely many sentences of the language a theorem that specifies the conditions under which an utterance of the sentence would be true and the conditions under which it would be false. So it's far from clear what kind of systematic, but evidently non-compositional, account of whatever sort of semantic properties expressions have if what I've said about the effect of vagueness on semantics is correct. So, until what needs to be explained—whatever exactly that is reckoned to be—is satisfactorily explained, perhaps the most that can be said for what *this essay* accomplishes is that it has displayed a new paradox about vagueness and meaning. I'll take that.

References

Buchanan, R. 2010. "A Puzzle about Meaning and Communication." *Noûs* 44 (2): 340–71. Buchanan, R. and Ostertag, G. 2005. "Has the Problem of Incompleteness Rested on a Mistake?" *Mind*, 114 (456): 889–913.

Chomsky, N. 1980. *Rules and Representations*. New York: Columbia University Press.

Frege, G. 1892. "Über Sinn und Bedeutung." *Zeitschrift für Philosophie und Philosophische Kritik* 100: 25–50.

Grice, H. P. 1957. "Meaning." *Philosophical Review* 66: 377–88.

Heck, R. 2002. "Do Demonstratives Have Senses?" *Philosopher's Imprint* 2: 1–33.

Lewis, D. 1969. *Convention: A Philosophical Study*. Cambridge: Harvard University Press.

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

_____ 1999. "Many, but Almost One." In Lewis, D. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press: 164–82.

McDowell, J. 1998. "*De Re* Senses." In McDowell, J. *Meaning, Knowledge, and Reality*. Cambridge: Harvard University Press: 214–27.

Neale, S. 2017). "Silent Reference." In Ostertag, G. (ed.). Meanings and Other Things: Themes from the Work of Stephen Schiffer. Oxford: Oxford University Press: 229–342.

_____ 1981. "Indexicals and the Theory of Reference." *Synthese* 49: 43–100.

_____ 1987. *Remnants of Meaning*. Cambridge: MIT Press.

_____ 2010. "Vague Properties." In Dietz, R. and Moruzzi, S. (eds.). *Cuts and Clouds: Vagueness, Its Nature, and Its Logic*. Oxford: Oxford University Press: 109–130.

_____ 2016. "Philosophical and Jurisprudential Issues of Vagueness." In Keil, G. and Poscher, R. (eds.). *Vagueness and Law: Philosophical and Legal Perspectives*. Oxford: Oxford University Press: 23–48.

_____ 2017. "Gricean Semantics and Reference." In Ostertag, G. (ed.). *Meanings and Other Things: Themes from the Work of Stephen Schiffer*. Oxford: Oxford University Press: 493–527.

Williamson, T. 1997. "Précis of *Vagueness*." *Philosophy and Phenomenological Research* 62 (4): 921–8.

Speaker Reference and Cognitive Architecture

DANIEL W. HARRIS*
Hunter College, New York, USA

Philosophers of language inspired by Grice have long sought to show how facts about reference boil down to facts about speakers' communicative intentions. I focus on a recent attempt by Stephen Neale (2016), who argues that referring with an expression requires having a special kind of communicative intention—one that involves representing an occurrence of the expression as standing in some particular relation to its referent. Neale raises a problem for this account: because some referring expressions are unpronounced, most language users don't realize they exist, and so seemingly don't have intentions about them. Neale suggests that we might solve this problem by supposing that speakers have nonconscious or "tacit" intentions. I argue that this solution can't work by arguing that our representations of unpronounced bits of language all occur within a modular component of the mind, and so we can't have intentions about them. From this line of thought, I draw several conclusions. (i) The semantic value of a referring expression is not its referent, but rather a piece of partial and defeasible evidence about what a speaker refers to when using it literally. (ii) There is no interesting sense in which speakers refer with expressions; referring expressions are used to give evidence about the sort of singular proposition one intends to communicate. (iii) The semantics–pragmatics interface is coincident with the interface between the language module and central cognition.

Keywords: Reference, compositional semantics, intentionalism, modularity, the semantics-pragmatics interface

1. *Intention-Based Semantics*

The aim of intention-based semantics (IBS) is to show how the concepts and claims that figure in our best semantic and pragmatic theories boil down to facts about the mental lives of human agents.

IBS has traditionally taken the form of claims like (1), which is Grice's mature explication of *utterer's occasion meaning* —what we now usually call 'speaker meaning' (Grice 1968, 1969).

(1)   "*U* meant something by uttering *x*" is true iff, for some audience *A*, *U* utereed *x* intending

   (i) *A* to produce a particular response *r*
   (ii) *A* to think (recognize) that *U* intends (1)
   (iii) *A* to fulfill (1) on the basis of his fulfillment of (2).

In keeping with the methodology of mid-Century analytic philosophy, Grice seems to have understood claims of this kind as conceptual analyses. But, following Schiffer (1982) and Neale (2016), we can modernize the project by recasting it in terms of metaphysical explanation. Let us therefore construe (1) as a grounding claim—a claim about what kinds of psychological facts are metaphysically sufficient for acts of speaker meaning.[1] And, moreover, let us suppose for present purposes that (1) is a *true* grounding claim. Why should we take an interest in it?

One answer is that (1) tells us something central about the nature of meaning and communication. To mean something by an utterance is to perform the speaker's half of an episode of communication. Communicative success further requires that the intended addressee recognize which kind of effect the speaker intends to have on them. By explicating speaker meaning as in (1), we learn that communication is a kind of mindreading—an application of our capacity to predict and explain agents' behavior by inferring their mental states. To mean something by an act is to use it to intentionally trigger and guide the mindreading capacity of an addressee, in part by revealing one's intention to do so.

Schiffer (1982) articulates broader ambitions for IBS. On his view, claims like (1) are crucial for the purposes of finding a place for meaning in the natural order. By showing how linguistic meaning boils down to speaker meaning, how speaker meaning boils down to human psychology, and, presumably, how human psychology boils down to physiology and ultimately physics, we would naturalize the subject matters of semantics and pragmatics, rendering them unspooky. In Schiffer's

---

[1] I take it that grounding is now a mature enough theoretical tool that I needn't spend time defending my use of it. The curious or sketpical reader can check out the following sources for elucidations and defenses of the concept: Fine (2012); Rosen (2010); Schaffer (2009; 2015). Although there is no lack of controversy about the nature of grounding, none of this controversy bears on my project here. I should also clarify that it would not matter for present purposes if we were to understand (1) as a claim about reduction, real definition, or supervenience rather than grounding.

view, this grand project hangs on the success of a string of claims like (1).[2]

Bracketing this grand naturalistic project, I think that claims like (1) can also offer us more modest and tractable payoffs. By revealing how our semantic and pragmatic capacities are grounded in independently understood psychological capacities, we open up new possibilities for *explaining* particular semantic and pragmatic phenomena, as opposed to merely describing and predicting them. Although this more piecemeal project is compatible with the grand naturalistic one, it also carries independent interest. Let me give two examples of what I mean before moving on to the business of this essay.

First, consider a pragmatic datum famously illustrated by Humpty Dumpty's botched attempt, in *Through the Looking Glass*, to mean 'there's a nice, knock-down argument' by uttering the words 'there's glory for you'. The lesson would seem to be that a speaker can't use any string of words to mean anything they want. But why not? What is it about the nature of speaker meaning that explains this fact? Some have been tempted to make the considerable leap from the falsity of Humpty Dumptyism to the truth of conventionalism, which is the idea that meaning something by an utterance is essentially a matter of conforming to linguistic conventions.[3] But conventionalism struggles to explain the many ways in which we communicate unconventionally—by speaking indirectly or non-literally, or by behaving passive-aggressively, for example. And these are cases that intentionalism has ample resources to explain. Each is a case of getting one's intentions recognized after all; it's just that we sometimes rely on unconventional evidence of our intentions to supplement, override, or take the place of linguistic evidence. What explains the apparently conventional constraints on speaker meaning, then? Intentionalists have replied that these constraints have nothing to do with conventions *per se*. Instead, they follow from independently motivated principles governing the interaction of beliefs and intentions in human minds. As most philosophers of action will tell you, it is either irrational or impossible to intend to do something that is ruled out by one's beliefs.[4] A rational agent who intends an addressee to recognize their intentions knows that they must provide evidence. A speaker can provide straightforwardly conventional evidence (as in the case of direct, literal speech), a mixture of conventional and overriding unconventional evidence (as in the case of indirect and nonliteral speech), or entirely unconventional evidence (as in fully non-conventional communication). This way of thinking about

[2] It should be noted that although Schiffer was initially optimistic about this project (Schiffer 1972; 1982), he eventually became disillusioned with it (Schiffer 1987).

[3] The classic defense of conventionalism is due to Searle (1965; 1969). A notable recent defense has been given by Lepore and Stone (2015).

[4] For variations on this principle, e.g. Bratman (1987); Broome (2013); Donnellan (1968); Grice (1971); Holton (2011); Neale (2004; 2016).

things rules out the kind of Humpty Dumptyism that we should want to rule out—namely, the idea that it is possible to mean anything by any utterance, irrespective of whether the speaker believes that their the utterance (together with whatever else is available) provides the intended addressee with adequate guidance. By linking speaker meaning to the speaker's intentions, we are thus able to explain a pragmatic datum in terms of what is independently known about human agency.

A second example of psychological explanation in IBS comes from my own recent work on the semantics of imperatives.[5] Here a central datum is that certain inference patterns involving imperative clauses strike us as valid. For example:

(2)   Buy me a drink!
      If you don't go to the bar, you can't buy me that drink.
      So, go to the bar!

One of the jobs of a semantic theory is to predict our intuitions about validity. But the usual strategy of taking validity to be a matter of truth preservation doesn't apply here, because imperatives aren't truth apt. IBS offers the crucial elements of an alternative strategy. The central ideas of this solution are plucked from Grice (1968; 1969), who argued that literal utterances of unembedded declaratives are intended to produce beliefs, that literal utterances of unembedded imperatives are intended to produce intentions to act, and that the meaning of a clause is a matter of the kind of effect that speakers use it to provoke. These ideas can be used to animate a formal-semantic theory on which the semantic values of declarative and imperative clauses, respectively, are the beliefs and intentions that they are characteristically used to produce, and complex, multi-clausal sentences encode more complex intentional mental states. On this view, we can predict that an inference will seem valid if a rational mind that exemplifies the semantic values of the premises also exemplifies the semantic value of the conclusion. These ideas allow us to recognize our intuitions about validity as the linguistic reflexes of underlying principles governing the structural rationality of beliefs and intentions. The inference pattern exemplified by (2), for example, reflects our sensitivity to the principle usually called *strict means-end coherence*:[6]

(3)   STRICT MEANS-END COHERENCE
      For any agent a and actions φ and ψ, a is irrational if a intends to ψ, believes that φing is necessary for ψing, but does not intend to φ.

---

[5] See my dissertation (Harris 2014) for the broad outline and my manuscript, 'Imperatives and Intention-Based Semantics' (Harris MSa), for the formal-semantic and foundational details.

[6] For defenses of some variations on (3), see Bratman (1987); Broome (2013); Holton (2011).

By marrying IBS to semantic theorizing in this way, my ambition is to move beyond the mere prediction of semantic data, and toward theoretically motivated explanations of them.

2. *Cognitive Architecture*

Intention-based semantics has traditionally trafficked in psychological states and processes that are wholly visible from the intentional stance. The key items in its explanatory toolkit are intentions, beliefs, and other posits of folk psychology. I think that we have reason to expand this toolkit. After all, the only rational communicators we know of are humans, and contemporary cognitive science has uncovered many phenomena in human minds that are not susceptible to folk-psychological understanding.

In particular, I think that we have good reasons to think of the mind as being carved up into one or more central cognitive systems and an array of peripheral *modules*.[7] An example of the kind of module I will discuss is the part of the human mind that is responsible for syntactic processing, which I will call the 'human sentence processing mechanism' (HSPM). The HSPM has several of the features of modules that interest me.

First, it is informationally encapsulated: in carrying out its operations, the HSPM has access only to a proprietary database of syntactic principles. It does *not* have access to information stored in central cognition, including the agent's conscious or unconscious beliefs, desires, or intentions. This is illustrated by the fact that the knowledge that one is about to perceive a garden-path sentence, such as 'the old man the boat', typically does not stop garden-path processing errors from taking place. A clear explanation is that garden-path sentences are unusual in that they violate the expectations embodied in the heuristics used by the HSPM to parse sentences' syntactic structures. Because the HSPM is modular, it lacks access to centrally available information, such as the belief that the sentence being read is a garden-path sentence. Central cognition, where this belief resides, simply isn't capable of intervening in order to avoid an error.[8]

---

[7] The classic defense of modularity is due to Fodor (1983). Recent defenses include Firestone and Scholl (2015) and Mandelbaum (2017). Note that I am not endorsing *massive modularity*—the view that even what Fodor took to be central processes can be decomposed into module-like subcomponents (Carruthers 2006; Sperber and Wilson 2002). However, what I say in this essay is compatible with massive modularity.

[8] Two caveats. First, some central cognitive processes can change what gets perceived. A much-discussed example is attention, which is at least partially under cognitive control, and which can alter the information available via perception, including linguistic perception. But recent proponents of modularity have argued that the effects of attention on perceptual input systems are limited to various forms of signal boosting and input selection, which do not amount to central information being used by a modular process, and so should not be considered a true violation

Second, the HSPM is centrally inaccessible, which is to say that its inner workings and the proprietary database of syntactic principles that guide it are not available to central cognition. Whereas encapsulation is a limit on the flow of information *into* a module, central inaccessibility is a limit on the flow of information *out of* a module. This feature is illustrated by the fact that we have no capacity to introspect either the processes by which the HSPM constructs syntactic representations or the syntactic principles on which it draws in doing so. This is what makes linguistics so difficult: we must laboriously reverse-engineer that to which we have no direct cognitive access.[9]

Modules have several other features that are worth noting. Modules are *domain specific*, in that they deal only with a proprietary genre of input. The HSPM deals only with *linguistic* representations, for example. Modules are *fast* and *mandatory*, in that they do their job quickly and in a way that is not subject to the agent's conscious will. We are powerless to avoid immediately perceiving as meaningful sentences of languages that we speak, for example, and this is because the HSPM is fast and mandatory. Moreover, the states and processes of the HSPM, like other denizens of modules but unlike central-cognitive states and processes, are ineligible for consciousness and are folk-psychologically intractable. Unlike the processes of central cognition, however, modular processes *are* susceptible to computational modeling. The HSPM is a prime example: syntacticians have made remarkable progress in modeling the proprietary database on which it draws, and psycholinguists have made remarkable progress in modeling the processes by which the HSPM acquires this database and deploys it in syntactic processing.

Modularity has been studied mainly as a property of perceptual *input* systems. But I will be concerned with modular *output* systems as well.[10] Take the HSPM: we need to build syntactic representations as part of the process of designing our own utterances, and not merely as part of the process of perceiving the utterances of others. To be sure,

---

of informational encapsulation (see, e.g., Firestone and Scholl 2015). Second, I am interested in modules qua *output* systems as well as input systems, and output systems are clearly responsive to central cognition in some respects, since that is where their inputs come from. It may help to clarify that the encapsulation of a modular system requires only that it it be insensitive to central representations *other than those that it is designed to take as inputs*. I will return to this issue below.

[9] Again, a caveat: the HSPM does send some information to central cognition, including the outputs of its perceptual processing and perhaps error messages when things go wrong. The point of inaccessibility is that these outputs, and not the various representations involved in producing them, are the only representations that bridge the gap between modules and central cognition.

[10] The standard focus on input systems begins with Fodor, who discusses them almost exclusively, although he does indicate that he is optimistic that much of what he says will also apply to "systems involved in the motor integration of such behaviors as speech and locomotion" (Fodor 1983: 42). For the idea that motor control is mediated by modular systems, see also Levy (2017); Stanley (2011).

speech production is not as well understood as the speech perception, but it is not hard to see that it bears many of the hallmarks of modular processes. Like speech perception, the syntactic processing of outputs is domain specific and fast. (It's less than clear what it would mean for it to be mandatory.) It is susceptible to computational modeling but opaque from the perspective of folk psychology and the speaker's own conscious thought. The details of syntactic processing, along with the database on which it draws, are just as centrally inaccessible on the way out as they are on the way in.

The question of encapsulation is complicated by the fact that speech production takes its marching orders from central cognition: my HSPM designs and outputs an utterance (with the help of other systems) as a result of my intention to speak or write. By concentrating, moreover, I can intentionally slow down the utterance-design process and consciously decide between different ways of formulating an utterance. I can ask myself whether it would be better to say 'driver' or 'chauffeur', for example, and I can even decide that the passive voice might sound better on a particular occasion.

However, there are severe limitations on the capacity of central cognition to intervene in the syntactic design process. Although this point about encapsulation is, strictly speaking, separate from the point that syntactic design is centrally inaccessible, the two points are easily understood together. Since central cognition lacks access to the principles governing syntactic construction, and to at least some of the concepts in terms of which these principles are framed, it can't very well intervene in a fine-grained way in that process. By way of example, consider an occasion on which I utter (4):

(4) Malik promised Kate to turn off the stove.

Our best syntactic theory tells us that the embedded infinitival clause in (4) has a phonologically null subject, PRO, which, since 'promise' is a subject-control verb, is bound by 'Malik'. So, the process leading up to my uttering (4) involved my HSPM representing it as having the following structure (and much more):

(5) $Malik_1$ promised Kate [PRO_1 to turn off the stove].

Let's assume that contemporary syntacticians are right about this, and that the HSPM of a competent English speaker who utters (4) represents the sentence being uttered as in (5). Clearly, this speaker would have no central-cognitive access to this representation. Indeed, most speakers would lack the conceptual resources to centrally represent sentences as in (5), since their central systems do not possess concepts of PRO, subject control, or coindexing, and are blissfully unaware that any part of them represents sentences as having properties like these. Even those of us who *are* aware of these facts did not become aware as a result of our central systems gaining access to our HSPMs, but rather as a result of a slow and grueling reverse-engineering project that has

taken decades and that remains incomplete. But since we have no central access to representations like (5), our central systems also have no way of intervening in the construction of such representations. A language user would be powerless to intentionally edit (5) so that it comes out as (6), for example—not just because they don't know how to centrally access representations of this kind, but because this sort of editing is not the kind of input that central cognition can send to the HSPM.

(6) Malik promised Kate$_1$ [PRO$_1$ to turn off the stove].

The bottom line is that 'promise' is a subject-control verb because the HSPM treats it as one, and our central system(s) simply have no say in this matter.

The syntactic processing of linguistic outputs therefore deserves to be thought of as encapsulated in the following sense: although this process takes inputs from central cognition, central cognition is powerless to intervene in its intermediate stages, or in ways that would require access to (and the ability to overrule) the HSPM's proprietary database. This leaves some interesting questions unanswered. Most pressingly: how rich are the inputs that central cognition sends to the modular components of utterance design? I cannot adequately address this issue here, but I will briefly return to it in §7.

The picture that emerges is one of mental subcomponents, including the HSPM and central cognition, that can transmit information to one another in only limited ways, and that would not be capable of handling many of one another's representations because they lack the informational and conceptual resources to do so.

These issues about cognitive architecture are relevant to IBS because IBS aims to reveal the psychological facts that ground our capacity to communicate with language, and there are good reasons to think that some of these facts concern modular input/output systems. Take the example I've just been discussing. My ability to produce and understand utterances of sentences like (4) depends on the fact that I have a properly functioning HSPM whose database includes principles governing control and binding. If this is so, then our strategy for implementing IBS will have to be constrained in some ways. We shouldn't assume that speakers' capacity to communicate with syntactically well-formed sentences is wholly grounded in facts about their beliefs and intentions, for example, since beliefs and intentions are denizens of central cognition. And likewise, any theory that takes the process of utterance design to be a rational, central-cognitive one that is wholly mediated by intentions and means-end reasoning will be flawed for the same reasons.

I have argued elsewhere that much of what semanticists study should likewise be thought of as a modular system—one that could be thought of as either a neighbor to or a subcomponent of the HSPM.[11]

---

[11] See the manuscript, 'Semantics without Semantic Content' (Harris MSb).

If so, then parallel issues apply to the ways in which IBS can fruit-fully study the psychological underpinnings of compositional seman-tics. This point will loom large in what is to follow, and I will return to it in some detail in §5. First, I need to say more about semantics itself, and about the role that reference is thought to play in semantics in particular.

3. *Reference*

Reference is widely thought to be among the central concepts of seman-tics and pragmatics, and there is a tradition within IBS of attempt-ing to show how facts about reference boil down to facts about human psychology.[12] A central thread running through this tradition is the idea that reference is, or is primarily, a thing that speakers do, not a relation borne by linguistic expressions (either types or tokens) to their referents (even relative to contexts). This fits with the broader strategy of explaining semantic facts in terms of facts about the actions or dispo-sitions of speakers, which are in turn explained in terms of facts about speakers' mental lives.

What makes reference such an indispensable concept in the first place, so that it deserves the full IBS treatment? As I see it, refer-ence standardly plays two important roles, one in semantics and one in pragmatics.

In standard semantic theories, reference is the relation that ties certain lexical items—type-*e* expressions, or referring expressions—to their compositional semantic values.[13] Since standard theories assume that the referents of these expressions function as inputs to semantic composition, semantics as we know it can't get off the ground unless reference supplies the raw materials. If we want to show how semantic facts, standardly understood, are grounded in facts about the psychol-ogy of rational communicators, then, these sorts of facts about refer-ence will have to be included.

---

[12] Aside from Neale (2004; 2016), who is my foil here, some other works in this tradition include Bach (1987; 1992); Bertolet (1987); Schiffer (1981); Stine (1978). An important precursor is Strawson (1950). Others who have given intention-based accounts of reference, though not explicitly in the IBS tradition, include Heim (2008); Kaplan (1989a); King (2013; 2014); Kripke (1977); Michaelson (2013).

[13] By "standard semantic theories", I mean those that build on the framework codified in the two most influential textbooks, Heim and Kratzer (1998) and von Fintel and Heim (2011). Of course, there are alternative frameworks, but most of the differences aren't ultimately relevant to the point of this essay. For example, in Jacobson's (2014) variable-free framework, the semantic value of 'he drinks' is the property of drinking, restricted to males. On Jacobson's view, it is a pragmatic matter for speaker and addressee to coordinate on a particular male and apply this property to them. But this is just to say that there is only speaker's reference on Jacobson's view—something that IBSers have long argued. It would therefore be easy to fit variable-free semantics into much of the dialectic that is to come.

There are some independent reasons to think that the concept of reference that is at work in semantics must be spelled out in terms of speakers' intentions. The point is perhaps clearest with respect to variables, including pronouns, such as 'she', 'it', and 'that'. Variables can occur either bound or free, and this is standardly accounted for by taking each occurrence to possess a numerical index and relativizing its semantic value to whichever assignment function is operative in the context. An assignment function is a mapping from numerical indices to elements in the domain of entities, D_e. The semantic values of unbound occurrences of pronouns are given by the following semantic clause.

(7) For any variable v, numerical index i, and assignment g, $[\![v_i]\!]^g = g(i)$

Thus we wind up with the following assignment-relative intension[as] the semantic value for 'he drinks', in which 'he' occurs unbound:[14]

(8) $[\![\text{He}_1 \text{ drinks}]\!]^g = \lambda w \, . \, g(1)$ drinks at w

Binding, on standard views, is understood as a compositional operation that reduces the assignment dependency of the resulting expression by λ-abstracting over all free variable occurrences with a given numerical index, turning them into argument positions in a complex predicate. By relativizing free variables' semantic values to assignment functions, we are therefore able to give a unified account of free and bound variables.

Where is *reference* in this picture? The standard answer is that unbound occurrences of variables are referring expressions, and the entities to which assignment functions map them are their referents. What refers to what is therefore a matter of the operative assignment. What determines which assignment is operative? Semanticists often dodge this question, or say something vague about "context" determining an assignment. For example, Heim & Kratzer say that "the physical and psychological circumstances that prevail when an LF is processed will (if the utterance is felicitous) determine an assignment to all the free variables occurring in this LF" (1998: 243). But most of those who have devoted serious thought to the question have argued that something to do with the speaker's intentions must be what does the job. For example, in more recent work, Heim says that "the relevant assignment is given by the utterance context and represents the speakers referential intentions" (2008). Kaplan (1989a), King (2013; 2014), and others have given similar, intentionalist accounts of how the referents of (at least some) unbound pronouns are fixed.

In addition to the foregoing semantic role for reference, it is also common to give reference a pragmatic role in explaining how speakers sometimes convey information about entities other than those to whom

---

[14] Most semanticists would say that this treatment applies only to deictic occurrences of pronouns, and distinguish, in addition to *deictic* and bound occurrences, unbound anaphoric occurrences. I will ignore this complication for now, since it's really just the deictic cases that interest me.

semantics tells us they should have referred. The most influential articulation of this idea is due to Kripke, who uses the following example (Kripke 1977: 263):

> Two people see Smith in the distance and mistake him for Jones. They have a brief colloquy: "What is Jones doing?" "Raking the leaves." "Jones," in the common language of both, is a name of Jones; it *never* names Smith. Yet, in some sense, on this occasion, clearly both participants in the dialogue have referred to Smith, and the second participant has said something true about the man he referred to if and only if Smith was raking the leaves (whether or not Jones was).

Kripke concludes that, in addition to semantic reference, we must also posit a notion of *speaker's reference*, which is also grounded in the intentions of the speaker, but which is less beholden to linguistic convention. Whereas semantically referring to Smith requires not only intending to say something about him, but also using an expression that linguistic convention provides for this purpose, an act of speaker's reference can break free of conventional constraints.

What should be the goal of IBS when it comes to to reference? If we take mainstream semantics and pragmatics at face value, we'll need to show how facts about both semantic reference and speaker reference are grounded in underlying facts about speakers' minds. One way to implement this strategy would be to distinguish *semantic reference* from *speaker reference* before explicating the former in terms of the latter and the latter in terms of underlying psychological concepts.[15] Such a strategy would be in keeping with the spirit of Grice's original articulation of IBS, in which utterance-type meaning (a.k.a. 'timeless meaning') is explicated in terms of speaker meaning, which in turn is explicated in terms of speakers' intentions.

This is not the main strategy that proponents of IBS have traditionally pursued, however, and so I will not focus on it here. Instead, IBSers have tended to argue that semantic reference, understood as a (context-relativized) relation borne by expressions to their referents, is not a genuine phenomenon, and that both the semantic and pragmatic roles of reference can be played by concepts of speaker reference. Put in terms of Strawson's (1950: 326) slogan, referring "is not something an expression does; it is something that some one can use an expression to do."

Although there are several versions of this idea, I will focus on the version developed by Neale (2016).[16] According to Neale, we need two

---

[15] As Neale (2016) points out, this is not Kripke's strategy in 'Speaker's Reference and Semantic Reference' (1977). Kripke instead does the reverse, explicating speaker reference in terms of semantic reference. King (2013; 2014) pursues a different strategy, putting both speaker reference and semantic reference on an explanatory par, while explicating both in terms of speakers' intentions (at least when it comes to context-sensitive referring expressions).

[16] Neale's theory of referring is a further development of Schiffer's (1981) theory. Some other accounts in a similar spirit have been developed by Bach (1987); Bertolet

concepts of reference. The first is what he calls 'speaker reference'

(SR)   **Speaker Reference**
       In φ-ing, S referred to o iff what S meant by φ-ing is an o-dependent proposition (a singular proposition that has o as a constituent).

As Neale points out, the idea here is to think of speaker reference as nothing more than a special case of speaker meaning—namely, the case in which what is meant is a singular, object-dependent proposition. If our theoretical repertoire already includes some concept of speaker meaning—for present purposes, it doesn't matter which version—and if we assume that humans sometimes communicate object-dependent information, then it follows without any further assumptions that Neale's notion of speaker reference is at least sometimes applicable.

One nice thing about (SR) is that it is highly versatile. In order to refer, on this view, there needn't be any linguistic expression *with which* one refers. As Neale (2016) argues at length, this is desirable because it allows us to make sense of the ways in which we can refer silently. In response to a question about what Smith is doing, one can say 'raking the leaves', thereby referring to Smith without uttering any expression with which one refers to him, for example. And, Neale argues, we should sometimes say that an agent has referred to someone or something even with an utterance that is neither linguistic or conventional in any way. Suppose that Malik visits Anne's apartment and finds that the place is a huge mess. "What happened!?", Malik asks. In response, Anne merely rolls her eyes and gives a furious look. Malik knows that only one thing can make Anne this angry—her good-for-nothing roommate, Chad—and so Malik correctly infers that, by her eye roll, Anne meant that Chad was responsible for the mess. According to Neale, Anne was referring to Chad, simply because her eye roll was a means of communicating a singular proposition about Chad. And I agree that this captures at least one sense in which communication sometimes involves referring. In particular, (SR) gives us a notion of reference that can play the pragmatic role, since it allows for the possibility of referring to someone without using an expression that, according to convention, can be used to refer to them.

But the concept of reference spelled out in (SR) can't play the *semantic* role that reference is usually thought to play. This is because (SR) gives us no resources to connect particular referring expressions to their referents on particular occasions—no way of linking referring expressions to their semantic values for the purposes of compositional semantics.

For this purpose, Neale identifies a second concept of referring:[17]

(RW)   **Referring-With**

(1987); Stine (1978). The central points of the present essay could be aimed at any of these views.

    [17] This idea is closely modeled on Schiffer's (1981) concept of referring-by. The two differ in some ways that are not relevant here.

> In uttering x, S referred to o with e, relative to its i-th occurrence in x, iff for some audience A and relation R, S intended A to recognize that R(e,i,o) and, at least partly on the basis of this, that S referred to o in uttering x.

Every instance of referring-with is also a case of speaker reference, but not vice versa. Referring-with requires more: it requires the speaker to have intentions about a particular occurrence of a particular expression, a particular referent for this occurrence, and a particular relation R that ties all three together. It is this added specificity that allows referring-with to play the semantic role by supplying occurrences of referring expressions with their semantic values.

Why shouldn't we just think of (RW) as Neale's definition of semantic reference? There are at least two good reasons. First, referring-with, so defined, is clearly something that speakers do with expressions, rather than something that expressions themselves do.[18] Second, and perhaps more importantly, referring-with needn't be mediated by linguistic convention in any way. This is because the relevant R-relation varies greatly between cases. It may sometimes be a conventional relation. For example: if I refer to Lincoln with 'Lincoln', it is likely that a conventional, lexically encoded relationship linking the name to the man is at least part of what plays the role of (RW)'s R-relation. But suppose that my friend and I see someone approaching in what looks like a pirate shirt, and I say 'what's the deal with Sinbad?'. In this case, I've referred to the approaching person with 'Sinbad', but not because 'Sinbad' conventionally refers to him. Here, I intend my friend to recognize that both the name 'Sinbad' and the approaching person's shirt are evocative of pirates, and it is this non-conventional relation that plays the role of the R-relation.

On Neale's view, then, referring expressions don't refer of their own accord. On a given occasion, it is the speaker's job to plug in their compositional semantic values by referring with them. Moreover, the semantic value of a referring expression on a particular occasion may be unconventional and ad hoc. Although the lexicon may contain the information that 'Lincoln' is sometimes used to refer to Lincoln, it is possible to override this sort of lexical guidance, using a name to refer to someone or something novel, in which case this novel entity serves as the name's semantic value.

What does this theory predict about Kripke's scenario, in which a speaker utters 'Jones' but conveys information about Smith? Neale doesn't say, but it seems to me that there are different options, depending on details of the case and background theoretical commitments about the metaphysics of intentions. A crucial aspect to this case is that the speaker falsely believes, of the man raking the leaves, that

---

[18] Interestingly, one could make the same case about Kripke's definition of semantic reference. However, my next point clearly does not apply to Kripke's definition.

he is Jones. So, it is at least initially plausible to attribute at least the following (RW)-instantiating intentions (and possibly even more) to the speaker on this occasion.

(RW1) S intended A to recognize that is-usually-called('Jones', 1, Jones) and, at least partly on the basis of this, that S referred to Jones in uttering 'what is Jones doing?'.

(RW2) S intended A to recognize that is-usually-called('Jones', 1, dthat[the man raking leaves]) and, at least partly on the basis of this, that S referred to dthat[the man raking leaves] in uttering 'what is Jones doing?'.[19]

(RW3) S intended A to recognize that S-uttered-x_n-while-looking-at-y('Jones', 1, Smith) and, at least partly on the basis of this, that S referred to Smithin uttering 'what is Jones doing?'.

(RW4) S intended A to recognize that S-uttered-x_n-while-looking-at-y('Jones', 1, Jones) and, at least partly on the basis of this, that S referred to Jones in uttering 'what is Jones doing?'.

One theoretical option for Neale is to say that the speaker has all of these referential intentions, and so refers to both Smith and Jones with 'Jones' in this case. On this view, the speaker's sentence has two different semantic values on this occasion. A second option would be to think of one of (RW1–4) as the speaker's only real intention, or as the *primary* or *governing* intention, to which the others are somehow subservient. On this view, it is the governing intention that matters, at least for the purpose of semantic reference.[20] A third option is to argue that in cases of this kind, when the speaker's intentions are incoherent or conflicting, something has gone so wrong that it does not make sense to say that any referring has occurred at all.[21] This idea is most plausible when we try to say what it would take for a hearer who is not confused about the identities of Smith and Jones to correctly interpret the speaker in this case. There seems to be no fully satisfying answer: although the hearer could diagnose the speaker's confusion, correct interpretation seems out of the question. I won't try to decide between these three ways of understanding Kripke's case here. I think that all three are worth seriously considering.

What is so attractive about Neale's account of referring-with?[22] From the point of view of IBS, the answer is clear: the view, if it works, gives us a concept of reference that does the work we need it to do in semantics and pragmatics, and that is wholly grounded in the same sorts

---

[19] Following Kaplan (1978; 1989b), I use ´dthat[the φ]` as a referring expression that refers to whatever is denoted by ´the φ`. For some complications about how to interpret 'dthat', see (Kaplan 1989a: 579–582).

[20] This is similar to a strategy that King (2013) uses to account for similar cases.

[21] For different versions of this idea, see Michaelson (2013) and Unnsteinsson (2016).

[22] And, by extension, what is so exciting about Schiffer's theory of referring-by, on which Neale's view is based?

of independently motivated cognitive resources—intentions, beliefs, etc.—that we know and love. The theory therefore delivers a seemingly crucial component of what IBS has promised us.

4. *The Aphonic-Intention Problem*

Unfortunately, there is a serious problem with this theory of reference as it stands, and with (RW) in particular. Neale calls this problem 'the aphonic-intention problem'.[23] It arises when we consider the full range of referring expressions that have been posited by contemporary philosophers and semanticists, some of which are *phonologically null* or, as Neale says, *aphonic*.

Consider (9), for example.

(9) It's raining.

Stanley and Szabó (2000) have influentially argued that sentences of this kind contain an aphonic variable (which I will write '*loc*'), which refers to a particular location on particular occasions of utterance. In uttering (9), one never says that it is raining, full stop, after all. Rather, one always says of some place that it is raining there. According to Stanley and Szabó, it is the aphonic referring expression *loc* that allows one to accomplish this act of referring, and so the LF of (9) includes *loc*, as in (10).

(10) It's raining loc_1.

A crucial part of Stanley and Szabó's argument for this claim—and, in particular for the claim that *loc* should be considered a *variable*—is that *loc* can seemingly be bound, as in (11).

(11) Everywhere I go, it's raining.

This sentence has a covarying reading, on which it means, 'everywhere I go, it's raining there'. According to Stanley and Szabó, the best explanation of this reading of (11) is that *loc* is bound by the adverbial quantifier, 'everywhere I go', as in (12).

(12) [Everywhere I go]$_1$ it's raining loc_1.

But if we need to posit *loc* in this case in order to explain covarying readings, theoretical economy recommends positing it as an unbound referring expression in (9) as well.

This argument is controversial, and various ways of resisting it have been proposed.[24] However, aphonic variables have been posited in analogous treatments of many other constructions, including quanti-

---

[23] This problem is a further development of Schiffer's (1992; 1994) "meaning-intention problem". But whereas Schiffer's problem deals with the implausibility of saying that language users have intentions about some of the things they are purported to refer *to*, Neale's problem deals with the implausibility of saying that language users regularly refer *with* certain expressions. I should say, at the outset, that I do not propose to deal with Schiffer's problem in this paper.

[24] See, for example, Neale (2007a;b; 2016); Recanati (2004; 2010).

fiers and definite descriptions (Stanley 2002; Stanley and Szabó 2000) , proper names (Fara 2015), adjectives (Kennedy 1999, 2007; Rett 2015), modals (Stone 1997), tense (Partee 1973), and many other expressions. And other aphonic expressions have been posited on syntactic grounds. These include PRO ("big pro"), *pro* ("little pro"), and *t* ("trace"), and the phonologically null pronouns in pro-drop languages, such as aphonic Italian subject expressions. Moreover, some of these, including Italian null subjects, can have unbound, referential occurrences. Thus we have a multitude of reasons for thinking that there are at least some aphonic variables, and I will assume that this is so.

The aphonic-intention problem arises because, according to Neale's view, referring expressions get their semantic values because speakers refer with them, and referring with an expression requires having intentions about it. But if many natural-language sentences contain aphonic referring expressions, this story becomes implausible, as there are excellent reasons to think that ordinary speakers do not know of the existence of these expressions. But how can someone have an intention about an expression about whose existence they are unaware? Without going so far as to endorse it, Neale articulates a reply to the aphonic-intention problem, which he calls the 'tacit-states reply'. This reply draws on Chomsky's (1980) idea that some of our knowledge of language is 'tacit', and Loar's (1981) idea that communicative intentions, and beliefs about them, are often unconscious, tacit mental states. In particular, according to the tacit-states reply, we must normally posit a tacit intention that instantiates (RW) whenever a speaker refers with an aphonic expression.

5. *The Aphonic-Intention Problem and Cognitive Architecture*

I don't think the tacit-states solution to the aphonic-intention problem will work. To be sure, I have no problem with the idea that some of our intentions are unconscious. Contemporary cognitive science has shown that our minds are replete with unconscious mental states, including beliefs and desires. It would be bizarre to think that intentions are somehow special in that they must be conscious. The claim that there are tacit intentions is somewhat more difficult to evaluate, since there is no settled understanding of what 'tacit' means. But, if we take it to mean that agents sometimes have intentions that they are unaware of themselves as having, and that they would not report themselves as having if asked, then I see no problem with this.

But deeper problems lurk. For example, there are some competent speakers who aren't merely *unaware* of the existence of aphonic expressions; because of their philosophical beliefs, they actively *deny* that aphonics exist. Let us stipulate, for the sake of argument, that they are mistaken in this belief. Still: their belief does not impair their ability to

discuss the weather, to use restricted quantifiers, or to speak Italian, and so it must not be interfering with their capacity to refer with aphonics. The tacit-intentions advocate is therefore not merely forced to say that agents are unaware of their tacit states and cannot report them; they must also say that these tacit states are unaffected by conflicting conscious beliefs and intentions. An agent who believes that aphonics don't exist, but who (tacitly) intends to refer with an aphonic, exemplifies a pattern of thought that intentionalists deem either impossible or irrational: they intend to φ despite believing that it is not possible for them to φ. This is a strange conclusion: although aphonic-deniers may be mistaken, it seems overly harsh to accuse them of irrationality. After all: they may have been led to their position by solid (if ultimately misleading) reasoning.

Similarly strange is the idea that many language users are able to have (tacit) intentions about aphonics despite apparently failing to possess some of the very concepts in which these intentions are framed. For example: most speakers seem to lack the concept of an aphonic, or of any of the particular aphonics that linguists and philosophers have posited. Aphonics like *loc*, PRO, domain-restriction variables, and null Italian subject expressions are all theoretical discoveries, and getting a lay speaker to have beliefs about them seemingly requires teaching them substantial amounts of linguistic theory, much as getting someone without scientific training to have thoughts about subatomic particles requires teaching them some physics. Lay speakers not only fail to have conscious thoughts about these aphonics; they seemingly lack the conceptual capacity to do so.

We can sum up the last two paragraphs by saying that, if speakers have tacit intentions about aphonics, then these intentions are both *inferentially* and *conceptually isolated* from their conscious mental states. This is not a feature that we normally expect from run-of-the-mill non-conscious mental states. And so, in positing tacit intentions about aphonics as part of our solution to the aphonic intention problem, we adopt a theory of tacitness that makes it something much more substantial than mere non-consciousness. We would need a theory of tacitness that allows tacit mental states to live a life of their own—one that is cut off from conscious mental states (and most non-conscious states as well).

Rather than developing an elaborate theory of tacitness, I think that we should seek an explanation in terms of modularity. If human minds contain representations of aphonic expressions, then these representations exist not in central cognition, but only inside of the language module.[25] The evidence for this hypothesis is relatively clear. The fact that language users' conscious thoughts are inferentially and conceptually isolated from their representations of aphonic expressions

---

[25] Of course, there may be various language modules or sumbmodules, but I will ignore this detail for now.

is clearly predicted, for example, since the language module is informationally encapsulated and centrally inaccessible. Because of encapsulation, a language user's denial that aphonics exist cannot interfere with the module's representation of aphonic expressions. And because the module and central cognition each possesses a proprietary database of information, framed at least partly in terms of proprietary conceptual resources that the other may lack, we should expect the module to be at least partly conceptually isolated from central cognition.

Indeed, many aspects of semantic representation smack of modularity. According to standard semantic theories, the principles governing semantic interpretation are framed in terms of concepts that most language users do not possess at the level of central cognition, including *functional application*, *assignment function*, *numerical index*, *semantic type*, and so on.

Take a simple sentence like (13):

(13) He drinks.

Standard compositional-semantic theories tell us that (13) has a reading on which processing it requires constructing a series of representations that includes (14).

(14) $[\![\text{he}_1 \text{ drinks}]\!]^{g,c} = \lambda w \ . \ g(1) \text{ drinks in } w$

If this representation were a belief, then we would have to attribute concepts like *assignment function* and *numerical index* to all competent language users. Semantic competence with aphonic variables includes these conceptual competences, but also competence with concepts of particular aphonic expressions like PRO and *loc*. We could say that such beliefs are tacit, but this would be no more satisfying than saying that the principles and conceptual competences governing the HSPM are tacit. Describing the situation in terms of modularity is preferable because modularity is posited explicitly in order to explain the ways in which certain mental states are informationally and conceptually isolated from central cognition, in just this way.[26]

A good working hypothesis is therefore that if being a competent language user entails representing sentences as in (14), these representations are not centrally accessible, and cannot be manipulated in any detail by central cognition. In particular, it seems clear that, in most language users, central cognition lacks the conceptual resources to work with with representations like (14). Indeed, we know about

---

[26] The idea that compositional semantics is the study of the inner-workings of a modular system would also explain a few other things. For example: semanticists have had a great deal of success with the project of computationally modeling humans' semantic competence—a hallmark of modular systems. Likewise, we tend to quickly and automatically experience linguistically formatted stimuli as meaningful, even when we believe them to have been the product of random forces. (The canned example is of stones on a beach, blown by a hurricane into a pattern resembling a sentence.) This suggests that semantic processing is, to a considerable extent, cognitively impenetrable. I consider some further arguments in Harris (MSb).

such representations only as the result of a grueling, decades-long re-verse-engineering project that is still unfinished. In other words: representations like (14), and all representations of aphonic expressions as well, are modular.

The idea that all of our representations of aphonics live inside of modules poses a devastating problem to Neale's theory. After all, intentions do not live inside of modules, but are paradigmatically central representations, along with beliefs, desires, and other posits of folk psychology. But from these facts, together with the fact that modules are informationally encapsulated and centrally inaccessible, it follows that there cannot be intentions to refer with aphonics that instantiate (RW). Such intentions would require violations of the informational and conceptual boundaries between central cognition and the language module. Indeed, given the role of assignment functions and numerical indices in semantic representations of referring expressions, it seems that nearly all instances of (RW) would involve violations of this kind.

6. *Compositional Semantics and Modularity*

I therefore think that Neale's tacit-states reply cannot save his theory of reference. In particular, I think that any theory of referring that requires speakers to have central representations of referring expressions in order to use them cannot be made to work. But without (RW), Neale has no notion of reference that can play the semantic role outlined in §3—no way, that is, of saying how the semantic values of referring expressions are fixed.

It might be tempting to try to solve this problem by finding some other kind of mental states that could play the role of fixing the referents of referring expressions—not intentions, but states of a kind that could reside wholly within the language module. However, I don't think this is a realistic option. My reason is that any such state would have to be capable of representing both referring expressions and their referents. The former is no problem, since the language faculty specializes in representing linguistic objects. But, in using a sentence containing a referring expression, humans can refer to *anything*, including things that the language module lacks the conceptual resources to represent, since representing such things requires background knowledge about the extra-linguistic world. In short: if the picture I have painted of human cognitive architecture is correct, then there is no place in the human mind for reference-fixing mental representations to reside.

I will therefore propose an alternative theory—one that I have defended elsewhere on related grounds.[27] The key to this view is that we must change our understanding of how compositional semantics works in such a way that there is no longer any semantic role for reference to play. According to standard theories, compositional semantic val-

---

[27] See Harris (MSb).

ues are *contents*, in Kaplan's (1989b) sense. The content of a referring expression is its referent, and the content of a (declarative) sentence is a proposition. Because some expressions have different contents on different occasions, the interpretation function must be relativized to contextual parameters, including assignment functions. Semantics, on this view, is the study of how the contents of complex expressions are determined as a function of the (possibly context-relativized) contents of their parts.

I think that we should abandon this view in favor of a different conception of semantics. On the account I prefer, semantics is the study of the semantic component of the language module. The job of this module is to encode and decode partial and defeasible perceptual evidence of speakers' communicative intentions. When someone speaks, their language module encodes in an utterance evidence of the general kind of thing that they mean, on the assumption that they are speaking literally and directly. When a hearer perceives an utterance, their language module decodes this evidence, which tells the hearer what sort of thing the speaker means, on the assumption that the speaker is speaking literally and directly. We should think of this evidence as *defeasible* because we sometimes mean things other than what the linguistic evidence would suggest. Suppose that I wryly say, 'Joel is a fine friend' in a situation in which it's obvious that I don't think Joel is a fine friend, in order to ironically implicate that Joel is actually a lousy friend. In this case, you have nonlinguistic evidence of my intentions that defeats the linguistic evidence I have given you. We should think of semantics as the study of *partial* evidence because the language module, being informationally encapsulated, has no access to information about the extralinguistic context, and so cannot decode information about the contents of expressions whose contents vary with context. Instead, what is encoded about such expressions—including many or perhaps all referring expressions—is information about the range of possible contents compatible with using them literally. On this view, we can think of the semantic value of a sentence as a property of propositions—namely, the property shared by all of the propositions that the sentence can be used to directly and literally mean on particular occasions of use. The semantic value of a referring expression can be thought of as a property shared by all of the things that a speaker can literally and directly refer to in uttering a sentence that includes the expression.

I am not the first person to defend this conception of semantics. In fact, many IBSers have articulated similar views, at least in the abstract. According to Bach, we should think of the semantic values of sentences not as propositions but as "propositional radicals", which are like propositions except that they do not fully specify the contents of context-sensitive expressions (Bach 1987: 2006). According to Sperber and Wilson (1995: 175) and Carston (2006: 633), semantic representations of sentences are not fully propositional, but are "schemas" that

must be supplemented with information available only to central cognition in order to arrive at a full representation of what is said. According to Schiffer (2003: §3.4), the semantic value of a sentence is its "character*", which is a partial specification of the content and illocutionary force of the sort of speech act that would normally be performed with the sentence. Even Neale has advocated a similar view. He argues that "a semantic theory for a language *L* will provide, for each sentence *X* of *L*, a *blueprint* for. . . what someone will be taken to be saying when using *X* to say something" (Neale 2005: 189). These blueprints, he is clear, do not fully specify contents, but provide hearers with only partial evidence of what a speaker has said.[28]

So there is a rich tradition within Intention-Based Semantics of denying that it is the job of semantics to fix the referents of context-sensitive expressions, including at least many referring expressions. I am only the latest participant in this tradition. However, until now, no advocate of this position has worked out the compositional-semantic details: no IBSer has given a precise account of what the semantic values of pronouns are, if not their (assignment-relative) referents, for example, and no IBSer has given a precise formal-semantic account of the kinds of non-propositional sentential semantic values they wish to posit.

I have attempted these tasks, and I will summarize my findings here.[29] The theory is designed to non-destructively extend standard semantic accounts, namely those of Heim and Kratzer (1998) and von Fintel and Heim (2011). So, let us begin with a standard sentential semantic value of the kind they posit—say, (14):

(14) $[\![\text{he}_1 \text{ drinks}]\!]^{g,c} = \lambda w \,.\, g(1) \text{ drinks in } w$

This is a proposition, a function from worlds to truth values. But because 'he' is an unbound variable, (14) is an *assignment-relativized* proposition; it specifies a proposition only relative to a given assignment function. Since, as Heim (2008) tells us, "the relevant assignment is given by the utterance context and represents the speakers referential intentions", and since the language module has no access to information about extra linguistic context (particularly when it comes

---

[28] For some other proposals that are similar in some ways, see Barwise and Perry (1983); von Fintel and Gillies (2008); Swanson (2016). For some recent defenses of the distinction between semantic values and the contents of speech acts—though for different reasons and with different implications—see Ninan (2010); Rabern (2012); Stanley (1997); Yalcin (2007).

[29] Note that there may be a much more elegant way to accomplish this task. For example, a more elegant implementation may be possible by drawing on resources from variable-free semantics (e.g. Jacobson 2014) or alternative semantics (e.g. Hamblin 1973; Rooth 1985; Kratzer and Shimoyama 2002; Alonso-Ovalle 2006), either of which is designed to deliver semantic values similar to those I discuss here. The semantics I sketch here is designed to non-destructively extend the most standard theories out there in order to show that my proposal does not depend on anything remotely exotic.

to others' mental states), (14) is not the sort of representation that the language module can work with.

What we need instead of (14) is a semantic value for 'he drinks' that needn't be relativized to an assignment or a context in any way—a context-free semantic value. I propose that this is the right sort of thing to play the role that we want:

(15) $[\![\text{he}_1 \text{ drinks}]\!] = \lambda p_{st} . (\exists x_e : x \text{ is male})(p = \lambda w_s . x \text{ drinks at } w)$

In English, (15) specifies the property possessed by every proposition p such that, for some male x, p is the proposition that x drinks. Suppose that you overheard someone say 'he drinks', but know nothing of their intentions, or of the context in which they were speaking. What could you know about what they had said? Assuming that they were speaking literally and directly, you would know that they had said, of some male, that he drinks. In other words, you would know that what they said is a proposition with the property picked out by (15). Situations of this kind—those in which you hear someone utter a sentence but know nothing about their intentions or the context—are useful think about because they approximate the position that your language module is *always* in. Although your language module is capable of decoding linguistic evidence, it is by nature unable to integrate the information it decodes with information about the extralinguistic context. And so, (15) is just the sort of semantic representation that we should expect an English speaker's language module to be capable of constructing. Notice, moreover, that (15) is not relativized to either an assignment or a context.

We can compositionally derive semantic values like (15) with minimal alterations to the standard semantic framework. We need only tweak the semantics of variables and add in a single composition rule. First, we must assign variables two kinds of lexical semantic value. First, we can have a single lexical entry that specifies the assignment-relative content of every variable.

(16) For any variable v and any assignment g, $[\![v_n]\!]^g = g(n)$

In the framework I am sketching, assignment relativity is present only at intermediate stages of semantic derivation. The semantic module uses variable assignments only to track binding and coreference relations as it makes its way up a tree. This relativization is always eliminated in the final representation of a sentential semantic value. It therefore doesn't matter what assignments map numerical indices to. For present purposes, we can stipulate that (unmodified) assignments always map each numerical index to itself, for example.

Next, we posit a second kind of lexical entry for each variable—one that specifies the sort of evidence that an unbound use of this variable gives about the speaker's communicative intentions. I will say that these lexical entries specify variables' 'constraint properties', and I will symbolize the constraint property of a variable v as $\mu(v)$. The constraint properties of 'he' and 'she' are given as follows:

(17) $\mu(\text{he}) = \lambda x_e \,.\, x$ is male

(18) $\mu(\text{she}) = \lambda x_e \,.\, x$ is female

In general, variables differ in meaning only insofar as they differ in their constraint properties.

Given this semantics for variables, how can we derive sentential semantic values like (15)? First, I make a syntactic assumption by positing bit of extra structure at the top of every sentence's phrase structure. Take S nodes to be the usual sentence nodes. Then I will assume that every sentence consists of an SA ("sentence abstract") node, with λp and S daughters, as follows:[30]

$$
\begin{array}{c}
\text{SA} \\
\diagup \quad \diagdown \\
\lambda p \qquad \text{S}
\end{array}
$$
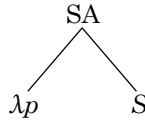
Given this syntactic assumption, what we want is a composition rule that transforms the semantic values of S nodes—i.e., assignment-relativized propositions—into the semantic values of SA nodes—i.e., unrelativized properties of propositions of the kind I described above. This principle should take us from $[\![\text{S}]\!]^g$ to $[\![\text{SA}]\!]$ in (19), for example:

(19)
$$
\begin{array}{c}
[\![\text{SA}]\!] \\
= \lambda p_{st} \,.\, (\exists x_e : x \text{ is male})(p = \lambda w_s \,.\, x \text{ drinks at } w) \\
\diagup \qquad\qquad\qquad \diagdown \\
\lambda p \qquad\qquad\qquad\qquad [\![\text{S}]\!]^g \\
\qquad\qquad\qquad\qquad = \lambda w \,.\, g(1) \text{ drinks in } w
\end{array}
$$

The following compositional principle does what we want, in the general case:[31]

(20) *Proposition Abstraction*

Let a be a branching node with daughters β and γ, where (a) β dominates only λp, and (b) γ contains unbound variables $u_i...v_n$. Then $[\![a]\!] = \lambda p_{st} \,.\, (\exists x^i : \mu(u)(x^i)) \,...\, (\exists x^n : \mu(v)(x^n))(p = [\![y]\!]^{g^{x^i/i...x^n/n}})$

The intuitive idea behind this principle is that it defines the last operation on sentences' semantic values before they are delivered to central cognition. All reference to assignment functions is eliminated, and so central cognition needn't have the concept of an assignment or a numerical index. Likewise, central cognition needn't be capable of forming representations of any particular lexical items, including aphonics. Instead, it need only be capable of forming representations of proposi-

[30] Although this syntactic assumption makes for a clean presentation, it is not essential to the view I am presenting. We could instead get what we need by assuming that the language module does one final transformation on its outputs before sending them to central cognition—a transformation that would be tantamount to the composition rule outlined below.

[31] Notation: subscripts on variables are numerical indices, as usual. Superscripts on variables aren't indices but merely devices for disambiguating variables.

tions, properties of propositions, properties of entities, and so on. In effect, (20) cleans away anything that central cognition lacks the conceptual resources to handle, and also relieves the language module of having to represent anything that *it* lacks the conceptual resources to handle.[32]

On this view, there need be no representation in the minds of speakers or hearers that links each referring expression to its referent. Instead, referring expressions play a characteristic role in the production and consumption of evidence about what speakers refer to in using these expressions. But we no longer have any need for the idea that speakers refer to things *with* these expressions, if this is to be understood in Neale's way, as involving mental representations that link each occurrence of a referring expression to its referent. We can dispense with (RW) altogether, and make do only with speaker reference (SR). Reference thus winds up playing a much less central role than is usually thought. It plays no role in semantics, and it plays a role in pragmatics only in the innocuous sense that speakers sometimes communicate object-dependent information in a variety of ways. Referring expressions semantically enable this sort of communication by giving speakers a way of encoding evidence of their object-dependent intentions, but that is all they do.

7. *The Semantics–Pragmatics Interface, Inbound and Outbound*

The picture I have sketched so far has some interesting consequences about the nature of the semantics–pragmatics interface. I have already indicated how I take the inbound semantics–pragmatics interface to work. The semantic module computes representations like (15)—context-free properties of propositions—and sends them upstairs to central cognition, whose task is to infer which proposition with this property (if any) is the one that the speaker meant.

But what about the outbound semantics–pragmatics interface? (Maybe we should call it "the pragmatics–semantics interface".) On the view I have sketched, this amounts to the following question: what kinds of instructions does the language module take from central cognition during the utterance-design process? This is a hard question to answer, and there is much less empirical evidence to guide us. For now, I will do no more than briefly point out some constraints on how the outgoing interface might work that follow from the picture I have sketched so far.

First, there are some constraints on what kind of instructions central cognition can send to the language module that arise from the con-

ceptual limitations of central cognition. Speakers can't intentionally meddle in the finer details of utterance design, because doing so would require having intentions about things that they do not, in general, have conceptual competence to centrally represent.

Second, there are some constraints on what kind of instructions central cognition can send to the language module that arise from the conceptual limitations of the language module. My language module isn't the sort of thing that can know that the person in front of me is named 'Tom', for example, because it lacks access to my beliefs about who is in front of me. If we think of propositions as intensions, or as structured Russellian complexes, then, we can't assume that the instructions sent to the language module consist of just a proposition, without any further instructions. Suppose, for example, that I am speaking to Tom, and I wish to tell him that he is silly. If propositions are structured Russellian complexes, then the content of what I say will be (21), and if propositions are intensions, then the content of what I say will be (22).

(21) ⟨Tom, silly⟩
(22) λw . Tom is silly at w

Given the situation I am in, this proposition (whichever kind of entity it is) is something I could say by uttering either (23) or (24).

(23) Tom is silly
(24) You are silly.

Given that I am addressing Tom, it would probably be much less confusing to utter (24) rather than (23) in order to say what I want to say. But this is not a choice that my language module can make, since it lacks access to my belief that I am currently addressing Tom. Since speakers can reliably and intentionally utter 'you' in order to signal that they are referring to their addressee, the fact that they are referring to their addressee must be included in the instructions they send to their language modules. There are many ways that this might be accomplished. But for now I will simply point out that one way it could be accomplished would be if the instructions sent by central cognition to the language module are just the same kinds of semantic values that I have posited above. For example, the semantic value of 'You are silly' could be represented as follows:[33]

(25) $[\![\text{you}_1 \text{ are silly}]\!] = \lambda p_{st}$. $(\exists x_e : x$ is the addressee$)(p = \lambda w_s$. x is silly at $w)$

On this view, outgoing instructions to the language module don't fully specify propositions, but only properties of propositions. Central cognition doesn't tell the language module what it intends to say, but only the general kind of thing it wants to say. Crucially, this "general kind of thing" is specified in a way that includes information about whether, for example, the proposition in question concerns the speaker's address-

---

[33] For more on how the semantic values of indexicals work in this framework, see Harris (MSb).

ee. Since this sort of semantic value is the same kind of thing that the hearer's language module decodes, there is a nice symmetry to this idea.

Although I can't claim to have defended the foregoing view here, I hope to have shown that it is attractive enough to be worthy of further thought.

8. *Speaker Reference without Semantic Reference?*

How does the view I have sketched make sense of Kripke's Smith–Jones case? Recall that in this case, the following dialogue takes place:

> (26) A: What is Jones doing?
>   B: Raking the leaves.

According to Kripke, "in some sense, on this occasion, clearly both participants in the dialogue referred to Smith" (Kripke 1977: 263). But it is also tempting to say that at least A, and possibly also B, has "in some sense" referred to Jones. And according to Kripke, we should resolve this dilemma by saying that A *speaker*-referred to Smith but *semantically* referred to Jones.

Like Neale, I have options about what to say about this case, and it is unclear which option is best. First we should keep in mind that both utterances must be explained as arising from a case of identity confusion: they believe, of the man raking the leaves (i.e., Smith), that he is Jones. Because of this state of confusion, it is unclear whether to attribute Smith-dependent communicative intentions to our speakers, or Jones-dependent communicative intentions, both, or neither. Take B's utterance. Which of the following intentions should be attributed to B?

> (27) B communicatively intends for A to believe that Jones is raking the leaves.
> (28) B communicatively intends for A to believe that Smith is raking the leaves.

One option is to say that there is no clear answer to the question of which of these was B's intention. As I said in §3, I am tempted to say that although it is possible for a properly informed hearer to diagnose B's confusion, there may not be a way of interpreting B such that genuine communication results. This line of thought would seem to recommend the conclusion that B has not genuinely referred to either Jones or Smith—B's thoughts are simply too muddled to do this kind of referential work.[34]

There is another option, which is to say that one of (27) or (28) is B's *real* intention—the one that really matters for communicative purposes—and that the other is unimportant. This becomes plausible if we flesh out Kripke's scenario in one of the following two ways. First, suppose that the main point of A's exchange with B is to discuss the man raking the leaves, whoever he is. We may suppose, for example, that

---

[34] For a defense of this idea, see Unnsteinsson (2016).

they are taking a walk around the neighborhood in order to see if *anyone* is raking leaves, and that they don't particularly care who. In this case, it is plausible that (28) is the intention that matters—the one that has to be interpreted in order for communication to succeed—and the fact that A used the word 'Jones' reveals merely incidental confusion in A's mind. On the other hand, suppose that A and B have taken a walk around their neighborhood in order to see what Jones is up to. In this case, since what they really care about is Jones, it becomes plausible to say that (27) is the intention that matters. On this view, the question of which intention matters in cases of confusion depends on broader facts about the goals and interests of those involved in the conversation.

Both of these ways of thinking about the case are plausible and worth pursuing, and I won't try to decide between them now. However, there is one remaining question about how I should make sense of Kripke's case. On the theory I have given, there is no such thing as semantic reference, and nothing that even plays the role it is purported to play. What, then, is the source of the intuition that A semantically refers to Jones in uttering 'What is Jones doing' ? My answer is that, in uttering 'Jones', A gives potentially misleading evidence about their communicative intention. Let us simplify the case somewhat, and suppose that A had said the following:

(29) A: Jones is raking the leaves.

Even if we ignore the above discussion and assume that A has speaker-referred to Smith in this case, why is it still tempting to say that A has semantically referred to Jones? The answer, I submit, is that A has given misleading evidence of their intentions. Specifically, I maintain that the semantic value of the sentence uttered by A is as follows:

(30) $[\![$ Jones is raking the leaves $]\!]$ =

$\lambda p_{st} . (\exists x_e : x \text{ is called Jones})(p = \lambda w s . x \text{ is raking the leaves at } w)$

If we suppose that what A meant is a Smith-dependent proposition, and that Smith is not normally called 'Jones', then the sentence uttered by A has a semantic value that gives B misleading evidence about A's intentions. This, in itself, isn't a problem: since semantic values encode *defeasible* evidence, B may be able to see past this evidence and recognize A's intention anyway. But still, it should not be surprising that we pay attention when speakers give misleading evidence of their intentions. It is this sort of misleading evidence, I submit, that causes us to posit a category of semantic reference when all we need is speaker reference.

9. *Conclusions*

The standard view is that it is expressions (or utterances of expressions) that refer, perhaps with some help from context. Intention-based semanticists have tended to follow Strawson (1950) in holding that expressions don't refer, though we can refer *with* them. Here I have

advocated a more radical departure from orthodoxy: expressions don't refer, and we don't refer with them either, but we do use them to give evidence of what we're referring to.

In arguing for this view, I have also advocated a broader account of the cognitive architecture underlying semantic and pragmatic competence. Pragmatics, on this view, is the study of a special kind of mindreading, wherein a speaker intentionally guides the mindreading capacity of their addressee, in part by revealing their intention to do so. Semantics, by contrast, is the study of a modular input/output system whose job is to encode and decode partial and defeasible perceptual evidence of speakers' communicative intentions. The semantic value of a sentence is just what we can know about what a speaker would be saying with it (if they were speaking literally), without knowing anything about their intentions or the context. On this view, the semantics–pragmatics interface turns out to coincide with the interface between the language module and central cognition.

References

Alonso-Ovalle, L. 2006. *Disjunction in Alternative Semantics*. PhD thesis. Amherst: University of Massachusetts.

Bach, K. 1987. *Thought and Reference*. Oxford: Oxford University Press.

Bach, K. 1992. "Intentions and demonstrations." *Analysis* 52: 140–146.

Bach, K. 2006. "The excluded middle: Semantic minimalism without minimal propositions." *Philosophy and Phenomenological Research* 73 (2): 435–442.

Barwise, J. and Perry, J. 1983. *Situations and Attitudes*. Cambridge: MIT Press.

Bertolet, R. 1987. "Speaker reference." *Philosophical Studies* 52 (2): 199–226.

Bratman, M. 1987. *Intention, plans, and practical reason*. Cambridge: Harvard University Press.

Broome, J. 2013. *Rationality Through Reasoning*. Oxford: Wiley-Blackwell.

Carruthers, P. 2006. *The Architecture of Mind: Massive Modularity and the Flexibility of Thought*. Oxford: Oxford University Press.

Carston, R. 2006. "Relevance theory and the saying/implying distinction". In Horn, L. R. and Ward, G. (eds.). *The Handbook of Pragmatics*. Oxford: Blackwell: 633–656.

Chomsky, N. 1980. *Rules and Representations*. New York: Columbia University Press.

Donnellan, K. S. 1968. "Putting humpty dumpty together again." *Philosophical Review* 77 (2): 203–215.

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

Fine, K. 2012. "A guide to ground." In Correia, F. and Schneider, B. (eds.). *Metaphysical Grounding: Understanding the Structure of Reality*. Cambridge: Cambridge University Press.

von Fintel, K. and Gillies, A. S. 2008. "Cia leaks." *Philosophical Review* 117 (1): 77–98.

von Fintel, K. and Heim, I. 2011. *Intensional Semantics*. Unpublished Lecture Notes, online at http://web.mit.edu/fintel/fintel-heimintensional.pdf, spring 2011 edition.

Firestone, C. and Scholl, B. J. 2015. "Cognition does not affect perception: Evaluating the evidence for 'top-down' effects." *Behavioral and Brain Sciences*, Target Article Under Comment as of 2015.

Fodor, J. 1983. *The Modularity of Mind*. Cambridge: MIT Press.

Grice, H. P. 1968. "Utterer's meaning, sentence meaning, and word-meaning." *Foundations of Language* 4 (3): 225–242.

Grice, H. P. 1969. "Utterer's meaning and intention." *The Philosophical Review* 78 (2): 147–177.

Grice, H. P. 1971. "Intention and uncertainty." *Proceedings of the British Academy* 57: 263–279.

Hamblin, C. L. 1973. "Questions in montague english." *Foundations of Language: International Journal* 10 (1): 41–53.

Harris, D. W. 2014. *Speech Act Theoretic Semantics*. PhD Dissertation, City University of New York Graduate Center.

Harris, D. W. MSa. "Imperatives and intention-based semantics." *Unpublished manuscript*.

Harris, D. W. MSb. "Semantics without semantic content." *Unpublished Manuscript*.

Heim, I. 2008. "Features on bound pronouns." In Harbour, D., Adger, D., and Bejar, S. (eds.). *Phi-Theory: Phi-Features across Modules and Interfaces*. Oxford: Oxford University Press.

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

Holton, R. 2011. *Willing, Wanting, Waiting*. Oxford: Oxford University Press.

Jacobson, P. 2014. *Compositional Semantics: An Introduction to the Syntax/Semantics Interface*. Oxford: Oxford University Press.

Kaplan, D. 1978. "Dthat." In Cole, P. (ed.). *Syntax and Semantics, Vol.9: Pragmatics*. New York: Academic Press.

Kaplan, D. 1989a. "Afterthoughts." In Almog, J., Perry, J., and Wettstein, H. (eds.). *Themes from Kaplan*. Oxford: Oxford University Press: 565–614.

Kaplan, D. 1989b. "Demonstratives." In Joseph Almog, J. P. and Wettstein, H. (eds.). *Themes from Kaplan*. Oxford: Oxford University Press: 481–563.

Kennedy, C. 1999. *Projecting the Adjective: The Syntax and Semantics of Gradability and Comparison*. Garland.

Kennedy, C. 2007. "Vagueness and grammar: The semantics of relative and absolute gradable adjectives." *Linguistics and Philosophy* 30: 1–45.

King, J. C. 2013. "Supplementives, the coordination account, and conflicting intentions. *Philosophical Perspectives* 27.

King, J. C. 2014. "Speaker intentions in context." *Noûs* 48 (2): 219–237.

Kratzer, A. 2002. "Indeterminate pronouns: The view from japanese." In Otsu, Y. (ed.). *Proceedings of the Third Tokyo Conference on Psycholinguistics*. Tokyo: Hituzi Syobo: 1–25.

Kripke, S. 1977. "Speaker's reference and semantic reference." In French, P. A., Jr., T. E. U., and Wettstein, H. K. (eds.). *Studies in the Philosophy of Language*. Minneapolis: University of Minnesota Press: 255–296.

Lepore, E. and Stone, M. 2015. *Imagination and Convention*. Oxford: Oxford University Press.

Levy, N. 2017. "Embodied savoir-faire: knowledge how requires motor representations." *Synthese* 195: 511–530.

Loar, B. 1981. *Mind and Meaning*. Cambridge: Cambridge University Press.

Mandelbaum, E. 2017. "Seeing and conceptualizing: Modularity and the shallow contents of perception." *Philosophy and Phenomenological Research*.

Michaelson, E. 2013. *This and That: On the Semantics and Pragmatics of Highly Context-Sensitive Terms*. PhD thesis. Los Angeles: University of California.

Neale, S. 2004. "This, that, and the other." In Bezuidenhout, A. and Reimer, M. (eds.). *Descriptions and Beyond*. Oxford: Oxford University Press: 68–182.

Neale, S. 2005. "Pragmatism and binding." In Szabó, Z. G. (ed.). *Semantics versus Pragmatics*. Oxford: Oxford University Press: 165–285.

Neale, S. 2007a. "Heavy hands, magic, and scene-reading traps." *European Journal of Analytic Philosophy* 3 (2): 77–132.

Neale, S. 2007b. "On location." In O'Rourke, M. and Washington, C. (eds.). *Situating Semantics: Essays on the Philosophy of John Perry*. Cambridge: MIT Press: 251–393.

Neale, S. 2016. "Silent reference." In Ostertag, G. (ed.). *Meanings and Other Things: Essays in Honor of Stephen Schiffer*. Oxford: Oxford University Press.

Ninan, D. 2010. "Semantics and the objects of assertion." *Linguistics and Philosophy* 33 (5): 355–380.

Partee, B. 1973. "Some structural analogies between tense and pronouns in English." *Journal of Philosophy* 70 (18): 601–609.

Rabern, B. 2012. "Against the identification of assertoric content with compositional value." *Synthese* 189 (1): 75–96.

Recanati, F. 2004. *Literal Meaning*. Cambridge: Cambridge University Press.

Recanati, F. 2010. *Truth-Conditional Pragmatics*. Oxford: Oxford University Press.

Rett, J. 2015. *The Semantics of Evaluativity*. Oxford: Oxford University Press.

Rooth, M. 1985. *Association with Focus*. PhD thesis. Amherst: University of Massachusetts.

Rosen, G. 2010. "Metaphysical dependence: Grounding and reduction." In Hale, B. and Hoffmann, A. (eds.). *Modality, Metaphysics, Logic, and Epistemology*. Oxford: Oxford University Press: 109–136.

Schaffer, J. 2009. "On what grounds what." In Manley, D., Chalmers, D., and Wasserman, R. (eds.). *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press: 347–383.

Schaffer, J. 2015. "Grounding in the image of causation." *Philosophical Studies* Online First.

Schiffer, S. 1972. *Meaning*. Oxford: Oxford University Press.

Schiffer, S. 1981. "Indexicals and the theory of reference." *Synthese* 49 (1): 43–100.

Schiffer, S. 1982. "Intention-based semantics." *Notre Dame Journal of Formal Logic* 23 (2): 119–156.

Schiffer, S. 1987. *Remnants of Meaning*. Cambridge: MIT Press.

Schiffer, S. 1992. "Belief ascription." *Journal of Philosophy* 89 (10): 492–521.

Schiffer, S. 1994. "A paradox of meaning." *Noûs* 28: 279–324.

Schiffer, S. 2003. *The Things We Mean*. Oxford: Oxford University Press.

Searle, J. 1965. "What is a speech act?" In Black, M. (ed.). *Philosophy in America*. London: Allen and Unwin: 221–239.

Searle, J. 1965. *Speech Acts*. London: Cambridge University Press.

Sperber, D. and Wilson, D. 1995. *Relevance: Communication and Cognition*. Oxford: Blackwell.

Sperber, D. and Wilson, D. 2002. "Pragmatics, modularity, and mindreading." *Mind and Language* 17: 3–23.

Stanley, J. 1997. "Rigidity and content." In Heck, R. (ed.). *Language, Thought, and Logic: Essays in honour of Michael Dummett*. Oxford: Oxford University Press: 131–156.

Stanley, J. 2002. "Nominal restriction." In Preyer, G. and Peter, G., editors, *Logical Form and Language*. Oxford: Oxford University Press.

Stanley, J. 2011. *Know How*. Oxford: Oxford University Press.

Stanley, J. and Szabó, Z. G. 2000. "On quantifier domain restriction." *Mind and Language* 15 (2–3): 219–261.

Stine, G. 1978. "Meaning other than what we say and referring." *Philosophical Studies* 33 (4): 319–337.

Stone, M. 1997. "The anaphoric parallel between modality and tense." *IRCS Technical Reports Series*: 1–44.

Strawson, P. F. 1950. "On referring." *Mind* 59 (235): 320–344.

Swanson, E. 2016. "The application of constraint semantics to the language of subjective uncertainty." *Journal of Philosophical Logic* 45 (2): 121–146.

Unnsteinsson, E. 2016. "Confusion is corruptive belief in false identity." *Canadian Journal of Philosophy* 46 (2): 204–227.

Yalcin, S. 2007. "Epistemic modals." *Mind* 116 (464): 983–1026.

Saying without Knowing What or How

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

In response to Stephen Neale (2016), I argue that aphonic expressions, such as PRO, are intentionally uttered by normal speakers of natural language, either by acts of omitting to say something explicitly, or by acts of giving phonetic realization to aphonics. I argue, also, that Gricean intention-based semantics should seek divorce from Cartesian assumptions of transparent access to propositional attitudes and, consequently, that Stephen Schiffer's so-called meaning-intention problem is not powerful enough to banish alleged cases of over-intellectualization in contemporary philosophy of language and mind.

Keywords: Implicit reference, propositional attitudes, omissions, intentionalism, transparency, phonology.

1. *Introduction*

Many linguists and philosophers of language believe there are linguistic expressions which are phonetically unrealized. Such expressions are syntactically real but lacking in phonetic and phonological properties. One of the most theoretically entrenched examples is (big) PRO which, according to current linguistics, occurs silently in sentences like

(1) $[_s[_{NP}$ Wanda$^1][_{VP}$ wants PRO$_1$ to win]]

and is anaphoric on its head NP. Clearly, the postulation of a silent expression like PRO raises all sorts of fascinating questions, some of which have been of particular interest to philosophers.

Stephen Neale, in 'Silent Reference' (2016), does an excellent job of bringing the questions and issues involved to the fore.[1] He is particularly concerned, as I see it, with showing that philosophers ought to be more careful and discerning in their use of this instrument in theorizing. Surely, his advice should be taken to heart. Philosophers need to consider when it is appropriate and plausible to posit phonetically unrealized expressions or syntax and when it is not. It may, for example, be all too tempting for, say, an epistemologist to say that speakers simply refer implicitly to epistemic standards whenever they use the word 'know.' But this raises all sorts of questions. How do they do so? Are they aware of doing it? And are they aware of doing it in the way the theory says they are?

However, in this paper, I argue that Neale's basic metaphysics is too restrictive to do justice to the theoretical options open to philosophers. He assumes, specifically, that it would be absurd to entertain the possibility of *uttering* phonetically null expressions. He also defines the class of aphonics of interest as expressions which, 'by their very nature,' lack phonological properties. I argue that these are mistakes and, further, that they are inconsistent with Neale's other commitments. And those other commitments are, by the look of it, more important. In the final section, I argue that Stephen Schiffer's so-called meaning-intention problem and Neale's related aphonic-intention problem are considerably less serious than they suggest. Borrowing a page or two from Peter Carruthers' (2011) work and from research in dual system psychology, I show that Schiffer and Neale make doubtful and controversial assumptions about the reliability or transparency of speakers' self-knowledge, making the meaning-intention problem far less effective in combating the alleged over-intellectualization of other theorists. Importantly, however, I argue that Gricean intention-based semantics can easily survive as a Cartesian divorcee, since meaning can still be determined by speakers' communicative intentions; they just don't necessarily have conscious awareness of the contents of those intentions.

2. *Mad Hatters, Cheshire Catters, and Troublemakers*

According to Neale, there is implicit reference and indirect reference. An object is referred to indirectly when a proposition which is merely implicated by a speaker has an object dependent truth condition. Implicit reference, however, occurs when a speaker expresses an object dependent proposition without there being any particular linguistic expression with which reference to the object is achieved. So, for example, if some philosophers are to be believed, and speakers can intend to refer to the location of the rain by merely uttering

(2) It's raining

---

[1] Page numbers in parentheses refer to Neale's paper unless indicated otherwise.

on a given occasion, then implicit reference is indeed ubiquitous in linguistic communication. In very general terms, there are two schools of thought on the nature of implicit reference: The Mad Hatters and The Cheshire Catters. The Hatters (for short) believe speakers can refer without there being anything at all with which they refer. They are as mad as a hatter, speak in riddles, expect their audience simply to work out what they intend and, just like the Mad Hatter, are punished *before* committing a crime rather than after (it's all in the pragmatics, you see). The Catters don't speak in riddles but they see non-existent objects everywhere, such as smiles and aphonic variables. In particular, they pretend to see these objects even when there is no theoretical need to do so.

Now, more precisely, the Catters are philosophers who wish to posit aphonic syntactic material in order to explain any plausible case of implicit reference. So, for example, just like linguists want to introduce the aphonic PRO in (1), Catters might propose to introduce an aphonic location variable in (2), which could give us (3)

(3) $[_S [_{NP} \text{It}][_{VP} \text{'s raining}[_{PP/AdvP} x]]]$

as a possible syntactic representation of (2).[2] On this model the aphonic variable could be occurring as an NP within a larger PP (substitutable for 'in Dublin') or as an AdvP (substitutable for 'here'). In this case, introduction of the variable is motivated, most obviously, by the claim that a location is necessary for an utterance of (2) to be evaluable for truth or falsity and, also, by the idea that the variable could be bound by an explicit quantifier, as in 'Everywhere I go, it's raining.' Mad Hatters like Neale and Schiffer, however, consider it much more important that ordinary speakers actually see themselves as having intentions to refer to a location when uttering a sentence like (2). More about that particular madness later (§4).

It seems like Neale wants in some sense to be both a Hatter and a Catter, so he takes on the role of Alice in 'Silent Reference,' trying to make sense of all the strange things in Wonderland. He tries to make the debate between Hatters and Catters more precise and starts by pointing out certain limitations of being a Catter. He points out that aphonic expressions like x in (3) and (unwritten) in (2), will have some rather strange features. First, they are proper parts of the sentences in which they occur but they never correspond to any part of any utterance of the sentence. This makes them very different from expressions like 'cake' and 'eejit.' Secondly, he argues, on this basis, that there can

---

[2] Note, however, that almost everything about (3) is controversial because, for one, expletive 'it' is here either a non-argument or quasi-argument. If it is construed as a non-argument—as in constructions like 'It seems that ...'—the gerundive 'raining' in (2) ought to be analyzed as CP with empty complementizer. It's also worth noting that many theorists would propose much more complicated analyses of a sentence like (2), involving multiple hidden variables—for time of utterance, the utterer, the world, etc.—I focus on the location variable here for simplification (see, e.g., Lewis 1970).

be no such thing as compositional semantics in which the meanings of parts of utterances compose to yield meanings for whole utterances, if one of the parts is supposed to be an utterance of an aphonic expression. A whole utterance of a sentence is a sequenced event which can be segmented into sub-events where each sub-event corresponds roughly to a word in the sentence. And, again, if 'eejit' is part of the sentence uttered, there will normally a be a roughly demarcated part of the utterance-event which corresponds to that word. No such utterance-parts will be found to correspond to PRO or x in (1) and (3). Therefore, compositional semantics cannot take as inputs the meanings of parts of *utterances*, if the semantic properties of aphonics are to play any role in composition. Composition must take as inputs the semantic properties of something other than utterances of expressions, it would seem; perhaps the expression-types themselves.

It would seem to follow, then, that one can't like utterance-based compositional semantics while being a Catter. But, of course, there are those who appear to do exactly that and we can call them Mad Catters (Stanley 2007 and Recanati 2010 are possible examples). After looking around for truthmakers to make Neale's two claims true, I realized I could find nothing but troublemakers. In what follows I discuss two such troublemakers before, in the next section, turning to more specific arguments against Neale's position. We should all be free to be Mad Catters when I'm done.

2.1. *Omissions*

Neale is rightly concerned with spelling out the nature of and connections between *words*, *sentences*, *utterances*, *propositions* and so on. Words are abstract artifacts created by linguistic communicative acts and sentences are then, presumably, abstract structures suitable to contain such artifacts in various syntactic arrangements. On Neale's view, utterances of sentences are events. Specifically, they are events whereby sentences are represented or, as he likes to put it, utterances are *proxies* for sentences. He makes the important point that the traditional distinction between expression-types and expression-tokens blurs and confounds these more fine-grained distinctions. There are not two fundamentally distinct kinds of linguistic expressions, i.e. types and tokens; there are, rather, expressions and various kinds of proxies for those expressions. A somewhat similar point has been made before (Searle 1978; Kaplan 1990) but the distinction still looms large in the literature and Neale makes particularly clear how detrimental to good theoretical sense it can be. Crudely put, utterances or inscriptions of sentences are not sentences any more than a picture of the Queen is the Queen.

Neale's discussion of aphonics would have been helped, though, by a more detailed examination of the kind of event an utterance or inscription is. As he is most certainly aware, utterances (let's ignore inscrip-

tions for now) are events under intentional description as their source is an intentional agent with goals, reasons, desires, beliefs and various cognitive and circumstantial limitations. In brief, they are intentional actions. Relatedly, the interpretation of an utterance by a normal hearer is geared towards the event as an intentional action: why did they choose those words? why are they saying what it seems like they're saying? Interpretation is geared towards reason-based explanations of intentional action. Linguistic interpretation—interpretation of speech acts—is just a special case of attempts at action understanding more generally. We automatically and effortlessly interpret human actions in terms of beliefs, desires, and intentions (e.g. Carston 2002: 42–44). Seeing someone walking repeatedly over some area in a field, their eyes moving quickly from one part of the grass below to another, I immediately assume they *want* to find something they lost, and that they *believe* it is there somewhere.

Already, this is a prima facie troublemaker for Neale's argument against Mad Catters. For ease of exposition, let's use 'action' for a complex intentional action and 'act' for any proper part of such a complex. What I mean by 'proper' part here is that the part is intentional just as much as the more comprehensive action of which it is a mere part. So, when I intentionally bake a cake, the act of breaking the eggs is an intentional proper part of the more comprehensive action. According to some philosophers, there are actions and acts that have no spatiotemporal properties at all. These are so-called acts of omission or refraining. Randolph Clarke (2010, 2012a, 2012b, 2014) argues, for example, that some omissions consist in the total absence of action relative to an agent, time, and location. Omissions can be unintentional or intentional; the latter he calls refraining. One further condition on refraining to V, for Clarke, is that there is some norm, standard, or ideal in place to the effect that one should V (2014: 29). There are others, however, who argue that refraining is always a type of action (e.g. Brand 1971; Fischer and Ravizza 1998). So, if an MP chooses to refrain from voting on a bill in parliament some particular bodily movement—or, even, the act of keeping still exactly then and there—must constitute the act of refraining at that time and place.

In fact, it doesn't matter what ontology of omitting or refraining we commit to, Neale's argument can only be saved if he can show that Mad Catters are, for some reason, not allowed to appeal to these notions in welding together aphonics and utterance-based compositional semantics. All parties to the debate agree that there is a sense in which people can intentionally omit to do something. Kent Bach (2010: 54–56) insists that, still, there is no sense in which refraining or omitting can count as actions or acts. But, as he realizes, omitting is not simply *not doing*. What counts as an omission, Bach agrees with Clarke, "is itself partly a normative matter" (Bach 2010: 54). So, whatever else it is, refraining from acting is part of folk psychology on all fours with acting, speaking,

expressing, and so on. That is to say, even if refraining to act is not really to act at all or consists in the absence of particular acts or actions, these non-acts can figure significantly in speech acts, their planning, and in the interpretation of speech acts. For our purposes, then, there is no harm in calling omission and refraining acts or actions. Just bear in mind that they could turn out to consist in the absence of an action or act on a given occasion—and so, strictly speaking, they are not actions or acts—or, alternatively, they correspond to something that was actually done. All we would need to do to accommodate Bach's insistence is to say that understanding intentional behavior in general is directed towards two kinds of objects; action and inaction.

To be clear, I am arguing that refraining to act on an occasion is in perfectly good standing, on anyone's account, when it comes to the automatic attribution of mental states to intentional agents in explaining their behavior. When I see someone accidentally drop a penny while walking in high-grown grass, stopping only for a fraction of a second to gaze down, I immediately assume they believe they lost a penny and that it is pointless to look for it. Arguably, Pennyless (their name) refrained from searching and I, watching, automatically explained this fact to myself by assuming various things about their mental state. If asked, Pennyless might confirm that searching for the coin would have been pointless, hence better to decide to do nothing at all. If I were Pennyless I would have done the same, I might think, and doing the same is the doing of nothing.

If refraining figures in action explanation generally, it also figures in utterance explanation in particular. As Clarke (2014: 32) points out, one of Strunk and White's famous dicta in *The Elements of Style* was "omit needless words." Clarke adds that, whenever one complies with this stylistic norm, one brings about the omission of words by the act of omitting their use. Furthermore, syntactic structure itself provides for a wealth of low-grade normative properties to capture the sense in which speakers refrain from uttering one thing in uttering another. So, for example, when I utter (2) while in Dublin it's clear to all that I should have added 'here' or 'in Dublin' if I wished to be more explicit, and if indeed my plan was to talk about the weather where I was located. There is a longer construction which I should have used in case I believed the context called for it. Let's say, then, that I refrain from saying explicitly where it rains in uttering (2). My refraining either consists in the absence of an act or it consists in some short-lived or instantaneous movement or other; quick breath, glance, gesture, whatever.

We have, then, candidate acts for being parts of utterances corresponding to aphonic parts of sentences. Utterances are actions which can, on occasion, be partly constituted by acts of refraining from saying something explicitly. Moreover, speakers can easily report on their acts of refraining after the fact. MPs may abstain from voting and report this by raising a hand or saying "I abstain". On some views, these lat-

ter actions would actually be spatiotemporally constitutive of the act of abstention but, as I have said, my argument doesn't require this assumption. I conclude that it makes perfectly good sense to say that, on occasion, speakers will intentionally perform the act of omitting to say explicitly. They do so, for instance, when they utter (2). This suffices to make trouble for Neale's argument. We have found a candidate to be the utterance-part corresponding to any plausible aphonic sentence-part. The candidate plays a significant role in speakers' capacities for mindreading, communicating, interpreting and explaining intentional action more generally. We could even imagine the communicative defects one would incur if one were, so to speak, omission-blind, and could only ever understand action, never inaction. Surely, this would be debilitating. And, finally, if there is an aphonic location-variable in sentence (2) it can correspond to the act of refraining from referring explicitly to a place.

2.2. *Gaps*

And what's the problem with instantaneous or durationless proper parts of utterances anyway? As research in phonetics and phonology shows, the correspondence relation between the abstract sentence or word and their audible utterance proxies is extremely complex and counterintuitive. This work has, for example, revealed what is sometimes called the 'lack of invariance problem,' namely that there is no one-to-one correspondence between acoustic signals and perceptual categorization into phonetic segments. In speech perception, different acoustic patterns are invariably perceived by hearers as a single pattern. Speech perception is 'categorical' on this way of thinking. The main reason for this is the phenomenon of coarticulation: the fact that discrete segments of speech are influenced acoustically by the immediately preceding or following sounds uttered. Take the articulation of the /p/ segments in 'pole' and 'peel.' The different positions of the lips, which is explained by the difference in the following vowels, creates differences in the acoustics. Normally, however, this difference is not reflected in the hearer's perception of the utterance (Unnsteinsson 2017).

Another counterintuitive feature, more relevant to our concerns, is that coarticulation occurs both within words and across words in flowing speech, resulting in the fact that, most of the time, gaps between words are not indicated by the continuous speech signal at all. Obviously, this is where inscription is usually very different. So, to take Neale's example, in uttering (4) the speaker sequentially produces five word-occurrences although the sentence contains only four words.

(4) The cat ate the mouse

Now, let's just assume the phoneticians and phonologists are right about all of this. This creates well-known problems about how knowledge of word boundaries is acquired so quickly and effortlessly by children. Yet

such knowledge can elude adults for a long time as well, for particular expressions (looking online for two seconds I found the example of 'housechablis' instead of 'house Chablis'). But this seems to be another troublemaker for Neale. When competent speakers utter (4) they will utter a sentence containing at least four word-gaps indicating that one word has stopped and another has begun. But the gaps in the sentence will almost never correspond to any identifiable gaps in any utterance of the sentence. The question is: how do speaker/hearers then learn where one word ends and another one begins? Presumably the answer has something to do with learning to identify similar acoustic patterns in different linguistic contexts. For example, competent speakers will also understand an utterance of, say, 'The mouse ate the cat.' But, of course, it will still be the case that almost no particular utterance of (4) contains parts corresponding to the word-gaps. So, competent speakers can utter a sentence with four gaps, understand effortlessly that the sentence has four gaps, without there being recognizable gaps in the utterance itself.

But the trouble doesn't start properly brewing until we reach the level of the phrase marker. If linguists are to be believed, sentence processing in ordinary speaker/hearers must involve the mental construction of an abstract syntactic structure. A fully developed human parser—the internal mechanism for processing sentences—at least assigns structures encoding various dependency relations between words and phrases to sentences encountered in speech and writing. In a recent book, David Pereplyotchik provides a wealth of arguments for the psychological reality of what he calls 'mental phrase markers.' The most telling arguments are based on results from brain-studies in neurolinguistics, so-called structural priming experiments, and on plausible explanations of garden-path effects. The data strongly suggest that there is an independent syntactic processing-stage which occurs before any semantic or pragmatic information is accessed. And this stage involves the construction of mental phrase markers identical to those developed by generative grammarians (Pereplyotchik 2017, Ch. 5).

Take structural priming, for instance (see Pickering and Ferreira 2008 for review). If a speaker encounters and parses a sentence with postulated phrase marker P then, the theory predicts, P has been activated in the speaker's mind. If P is activated it should remain so for a while and should show up in other mental processes. Experiments confirm that there are strong priming effects of this sort. So, if P is primed in sentence perception it becomes much more likely that a P-sentence is produced later, even if semantically equivalent sentences with other phrase markers are equally or more salient in the context. The human parser, it seems, automatically assigns phrase constituency structure to sentences.

It's important to note that, in spite of this, many theorists would argue either (i) that speakers have no knowledge or beliefs about mental

phrase markers; even that they aren't mentally represented, or (ii) that any such knowledge or belief is tacit, subpersonal, subdoxastic, or inaccessible to consciousness.[3] Cognition involving mental phrase markers, on most accounts, does not reach personal-level explanation. This contrasts with speakers' beliefs about what they intend to say or refer to on a given occasion of utterance. For normal humans, it seems trivially true that they know what they mean by uttering something. They seem, at least, to know some substantial part of it, as manifested in the capacity to repeat, clarify, or paraphrase what was meant (although I will criticize this alleged truism in §4). Still, supposing that competent speakers stand in some cognitive relation to mental phrase markers, and that the correct theory of this relation either falls into category (i) or (ii), sentences will have parts with no corresponding parts in the utterance. The most extreme views in category (i) will surely deny the very existence of mental phrase markers, but we can set them aside for the moment (see Collins (2007) for a moderate (i)-type view).

Assume, then, that speaker/hearers have tacit knowledge, at least, of the immediate constituents of a sentence. They automatically process sentences in terms of mental phrase markers. So, speakers tacitly know about NPs and VPs and constituency-boundaries are even posited as parts of the abstract syntactic structure of a given sentence. But what, if anything, is the part of an utterance of a sentence which corresponds to the part of the sentence-cum-phrase-marker that distinguishes the VP and the NP? Given that there are verb-subject-object languages, such as Irish, where the VP is split by the NP in normal word order (Irish is a VSO language), it's unclear how this question could be answered directly. The question falsely presumes that sentence-parts and utterance-parts that go proxy for sentences stand in simple, isomorphic mapping relations. Syntactic theory shows that a lot of material is properly said to be part of the abstract sentence, while having no obvious counterparts in utterances or inscriptions. So, the fact that some words have this feature as well should not be objectionable as such.

3. *Two Arguments for Uttering Aphonics*

Neale reports that when he talks of aphonics he is particularly concerned with "… individual expressions that are unpronounced and unheard *by their nature*, expressions that intrinsically lack phonological features or instructions for pronunciation" (236, italics in original). The idea seems to be that positing aphonics wouldn't be theoretically exciting unless the lack of phonetic and phonological properties is essential to the posited expression. Most theorists allow for aphonics in the less substantial sense in which they are actually phonic expressions that

---

[3] See Devitt (2006) for an example of the first kind of view and Dwyer and Pietroski (1996) for an example of the second.

happen to be omitted on a given occasion. In VP-ellipsis, for example, it is important that what is elided be identical to the antecedent verb—which can either precede or follow the ellipsis—as in 'Sally drove to England and Joe [drove] to Scotland.' Neale is interested, it seems, in expressions partly individuated by their lack of phonology, so that trying to utter them would always result in the production of some distinct expression.

This is unfortunate for a number of reasons, the most important being the fact that there is no such thing as an intrinsically aphonic expression. Any expression can be uttered, even if it happens to lack phonetic features or instructions for pronunciation. What's more, Neale explicitly recognizes this elsewhere in his paper.

3.1. *Uttering*

Neale provides a thoroughly intention-based theory of utterance identity. The question at issue is a metaphysical one: In virtue of what facts is a given utterance and acoustic proxy for a given word? Very roughly and relative to "a few reasonable assumptions," Neale writes that "an utterance u produced by S on a given occasion is an utterance of expression e iff S intended u to be an utterance of e" (265). On the face of it, this view of utterance identity appears to be incompatible with expressions which are aphonic 'by their nature.' We can plug any allegedly aphonic expression into Neale's biconditional and get, as a result, a speaker's utterance of that aphonic expression. So, for example, I can intend to utter the aphonic expression 'PRO' by articulating the sound /pro/ on a given occasion.

What's more, uttering aphonics is in some sense made easier by their lack of phonetic features. There definitely is already a standardized way of uttering the aphonic expression 'PRO,' at least within linguistics. But, for other less entrenched cases, such as location variables, it seems like we can choose any phonemic pattern that would do the job in the context at hand. Or, in lieu of that, one could utter the expression by omission, as discussed above. This fits with a theory of speech errors I have defended elsewhere, which also incorporates a thoroughly intention-based view of utterance identity (Unnsteinsson 2017). Very roughly, the idea is that when a speaker has expression e_1 as their target but accidentally produces some other expression e_2, the uttered expression will be a misarticulation of the target expression. It was an odd way, and accidentally so at that, of uttering the expression which was the speaker's intended target. Thus, I could intend to utter 'Obama' but accidentally utter 'Osama.' On this theory of speech error, I will have pronounced 'Obama' as 'Osama' on that occasion.

So, it seems like we have two options, neither of which allows for the possibility of intrinsically aphonic expressions. First, we could say that any utterance u where the speaker intends to utter the aphonic e by making u is such that e was in fact successfully uttered. Since there are

no instructions for pronunciation one really cannot go wrong and any-thing one does—provided one has *e* as one's target in uttering *u*—will count as an acoustic proxy for the aphonic. Secondly, we could say that the correct pronunciation of an aphonic expression is in fact silence. The instructions for pronunciation indicate that the expression ought to be uttered by making no sound at all. In effect, this is taking lack of phonetic features to be a kind of phonetic feature. But we don't need to choose between the two options, for both support the same conclusion, as noted before. If we take the second option, all utterances of PRO, for example, where the speaker intends to utter PRO by making the sound PRO (or any other sound) will simply count as a misarticulation of the aphonic. Importantly, however, and this is clearly part of the in-tention-based approach to utterance identity, any such misarticulation will still constitute an utterance of the expression. It follows, then, that intentionalists must believe that aphonics—though they're usually not heard in utterances—can very well be uttered and heard.[4]

But what on earth would it be like for an ordinary speaker to have an aphonic expression as their target? And how could some sound they emit constitute an utterance of that aphonic target? This is part of Neale's aphonic-intention problem, which I will discuss more directly below. Let's consider these questions naively first. It would seem like some attempts by ordinary speakers to make themselves absolutely clear because of prior misunderstanding might be classified as (tem-porally extended) utterances of an aphonic like PRO. Say I'm planning a road trip with you and Siobhan and we're deciding who shall drive. We've been uttering sentences like 'I want you to drive' and 'Siobhan wants me to drive,' so this syntactic structure is primed and we are thus a bit more likely to misidentify similar structures. Then I say,

(5) You or Jane wants to drive,

trying to transfer the responsibility for driving over to you or Jane. Primed for misunderstanding, you ask: 'We want *who* to drive?'.[5] When I respond Jane is no longer present. Somewhere along the way, my utterance-plan goes badly awry, but what comes out of my mouth is something like the following:

(6) *I said you or Jane wants *yours-her-self* to drive

Now, of course it is far from obvious what exactly we should say about this strange case. Perhaps I had the phonic expression 'yourself' first as a target and then, thinking I could fix the error, I had the phonic expression 'herself' as a target. So, didn't I just misarticulate those expressions? Probably, yes. But let's assume PRO exists and is really an aphonic pronoun controlled, in a case like (5), by the subject of the

---

[4] For a very different point of view on this, see Hawthorne and Lepore (2011: 460–465) and Lepore and Stone (2015: 217–220).

[5] If the misunderstanding involved here sounds implausible, just imagine this all happening over walkie-talkie in a movie from the 80s.

matrix verb, i.e. 'You or Jane.' Add to that the idea that PRO inherits
the reference from its antecedent. It's not crazy to suppose, then, that
when I made the error and uttered /yours-her-self/ my target was an
expression with exactly those features. I just couldn't find any phonic
expression in my mental lexicon which corresponded well to those fea-
tures. The problem is that the subject, 'You or Jane,' can easily control
aphonic PRO, but it can become awkward with an overt reflexive pro-
noun as in (6).

Still, I don't want the argument to rest entirely on the plausibil-
ity of this kind of case as it may be judged a bit far-fetched. However,
even if ordinary speakers never have an intention to, as we might say,
give phonetic realization to an aphonic, it is arguable that experts both
could and routinely have such intentions. When linguists or philoso-
phers utter or inscribe 'Wanda wants PRO to win' or 'It's raining in *x*'
one possible description of what they're doing is that they are uttering
or inscribing the posited aphonic expression. To support this, I see no
logical or metaphysical impossibility in the idea that after a few de-
cades the use of PRO would catch on in the general population. This is,
of course, terribly unlikely, and even more so in the cases which are of
particular interest to philosophers, like the ideas of aphonic location-
variables or modes of presentation.

There is, however, an obvious objection which is implicit in Neale's
own paper. As he notes, when linguists write PRO in a sentence, it's
part of their structural description of the sentence (243). So, one could
say, rather than uttering or inscribing the aphonic expression, what
linguists are doing is describing, or perhaps simply naming, the expres-
sion. Indeed, since the postulated expression has no phonic properties
it stands to reason that the expression is either named or described, not
uttered or inscribed, and that this is what the experts intend. I want
to fully acknowledge the strength of this point but, it's equally clear,
it still doesn't amount to showing that aphonic expressions like PRO or
variables for locations are aphonic 'by their nature.' If intentionalism
about utterance identity is assumed, experts can certainly utter these
expressions if they want to.

In his discussion, Neale introduces PRO as "silent *self*" which corre-
sponds to what he calls "stilted-'self'" but he makes clear that he thinks
they must be different expressions. And it's clear why he thinks this:
one is by its nature aphonic and the other is an odd or stilted extension
of the phonic word 'self' into a set of unfamiliar syntactic distributions,
namely exactly the distribution of PRO or silent *self*. It follows from
Neale's assumptions—although, as already noted, it's not compatible
with his notion of utterance identity—that silent *self* and stilted-'self'
are different. The former is not stilted and the latter is not silent (243).
But why suppose that this is a robust criterion for individuating words?
Well, we shouldn't suppose so. Before arguing for this claim, and re-
sponding properly to the objection from structural descriptions, we

need to go through the second argument for the claim that there are no intrinsically aphonic expressions.

3.2. *Uttering What?*

Let's agree with Neale that words are abstract artifacts, along with things like laws and conventions. Agree also that word-proxies or to-kenings of words are not words (see his Section 6). But can the nature of words as abstract artifacts be described in more detail? It would appear so. Wolfram Hinzen and Michelle Sheehan argue, for example, that there are four important notions of 'word' all of which play significant roles in current linguistics.[6]

> First, there is the notion of the word as a *prosodic* unit (the 'phonological word'); then there is the notion of the semantic word or *lexeme*, which is the word understood as an abstract vocabulary item with a given meaning that can take different forms, such as the verb RUN, which can take the forms *runs*, *ran*, *run*, etc. Even more abstract is the notion of a lexical *root*, which involves a semantic core possibly shared across lexemes of different categories, e.g. the root √RUN as involved in the verb *run* and the homophonous noun *run*, which occurs in the expressions a *run*, *many runs*, *running*, *Mary runs*, etc. Finally, there is the *grammatical* word: the word as a *morphosyntactic* unit or as functioning in a sentence context. (38, italics in original)

Neale agrees with Kaplan (1990, 2011) and others that the first unit on this list is not what we're looking for when asking about the metaphysics of words. Of course, there exists such a prosodic unit, but it's clear that there are psychologically real and important distinctions between phonologically and phonetically identical words. More importantly still, prosodic units can change so dramatically over time that their individuation is problematic. Kaplan (2011) and Hawthorne and Lepore (2011) argue for something like a lexeme-based theory of word identity. On their account, then, words are essentially objects with certain syntactic markers or properties as well as semantic ones: they are verbs or nouns for instance. As Hinzen and Sheehan (ibid.) point out, the category of lexeme is almost identical to what many philosophers call 'concepts.' Then there is also the more abstract notion of lexical root, which is shared by different lexemes. Finally, there are morphosyntactic units, characterized, for example, by the position the unit occupies in a phrase-structure tree or the manner in which it interacts with various affixes.

This provides a wealth of possibilities for how philosophers could define words. The correct metaphysics of words might incorporate any combination of these four properties, and of course, which ones are appropriated may depend on the theoretical purposes at hand. Most theorists seem agreed that, when individuating words for the purposes of describing the items stored in the mental lexicon, prosodic units are

---

[6] Thanks to James Miller for alerting me to this passage (Miller, 'The Metaphysics of Words,' unpublished).

not terribly important. There certainly are prosodic units but they are non-basic. It's reasonable to suppose, then, that a metaphysical theory of word identity will always incorporate at least one of three properties: (i) being lexemic, (ii) having a lexical root, and (iii) having a morphosyntactic profile. Words don't need phonetic realization—because we are assuming that there are aphonics—and whatever phonological properties they have can change dramatically, fluctuate or even disappear. Semantic properties of words allow for similarly dramatic fluctuations over time.

But here Neale's point about the difference between silent *self* and stilted-'self' may seem relevant: Wouldn't aphonic and phonic units always constitute distinct expressions? No, they would merely be distinct prosodic units, because all of the other properties could still remain intact. Again, as a comparison, does an expression with semantic properties become a different expression if it evolves into an expletive, 'non-semantic' unit? For instance, is expletive 'it' distinct from 'it' occurring as argument? The answer to these two questions would depend on whether or not one endorses the lexeme-based view of word identity. But the answer to the former, it seems, has to be a direct 'No.' Phonetic and phonological properties are superficial, non-basic features of words. Therefore, there is no such thing as an intrinsically aphonic expression.

Now we are also in a position to respond to the objection mentioned near the end of Section 3.1 above. According to the objection, when experts appear to be uttering aphonic expressions like PRO, silent-*self*, or variables for location, what they are really doing is providing structural descriptions or merely naming the aphonic item. Well, if both of my arguments are sound, this is seen to be unduly ad hoc. If phonological properties are superficial and utterance identity is determined by speakers' intentions, nothing hinders experts in intending the production of a particular sound as the utterance of some postulated aphonic expression.[7] Before, I gave reasons to think that if such expressions exist at all, the omission of any phonic counterpart in an utterance may count as an act of uttering an aphonic. I believe it is difficult to find credible cases of non-expert speakers in fact having intentions to utter an aphonic *by producing some sound or other*. Doing so, however, is open to the experts themselves. All I needed to show is that no word— apart from mere 'phonological words' of course—is aphonic by its very nature and, so, it is fine if it turns out that experts normally consider themselves only to be describing or naming aphonics, rather than actually giving voice to them. But when they in fact intend to utter the aphonic by producing some sound or other, the identity of the word in question isn't suddenly altered.

---

[7] There is one other possibility, however. One might simply argue that aphonics aren't expressions at all, that they are much more like phrase structure trees, Case, and other parts of the syntactic description of sentences. I find this possibility appealing, but won't address it here, as it would take us too far into different territory.

4. *What About the Meaning-Intention Problem?*

According to Schiffer and Neale, there is a significant difference between theories on which speakers have intentions to refer implicitly to something like a rain-location and ones on which they have intentions to refer implicitly to modes of presentation or epistemic standards. When it comes to rain, speakers seem immediately and effortlessly aware that they in fact intended to refer to a location even if they omitted any expression whose function is specifically to enable such reference. Ordinarily, when speakers say that it's raining, they'll have no trouble answering the question 'Where is it raining?' if it is somehow unclear in the context. So, it may seem, we have good reason to suppose that speakers actually have intentions to implicitly refer.

Modes of presentation appear to be different. Many philosophers wish to posit modes of presentation to solve puzzles about singular reference. According to one influential theory of this sort, when speakers attribute beliefs to others with sentences like (7),

(7) Bianca believes that the hippopotamus is sleeping,

they really express the proposition that Bianca believes, under some mode of presentation, that the hippopotamus is sleeping. So, uttering a sentence like (7) involves implicit reference to something called a mode of presentation (see Neale 2016, §14.4 for details). But, on the face of it, speakers are not aware of intending anything of the sort. If the speaker were asked to specify which mode of presentation they had in mind in uttering (7) the question would usually not be understood. Maybe there are more intuitive ways to get at the question, which would better track what theorists have in mind by positing these modes, but still, the difference between this case and the first one will remain. Normal speakers are completely aware that they implicitly refer to locations all the time, but, if they do indeed refer to modes of presentation—or epistemic standards, according to Schiffer—they haven't the faintest idea that that's what they're doing. It follows, then, that only cases of the first kind are compatible with the assumption that speakers have transparent, privileged access to what they consciously mean, intend, and believe in uttering something. And it is reasonable, Schiffer contends, to think that such access is "part of a normal person's functional architecture" (1992: 515; Neale 2016: 320).

Neale argues that aphonic reference presents an even deeper problem (2016, §14.7). Indeed, if there are theorists who hold that speakers refer to modes of presentation *with* aphonic expressions, it follows that they refer to things they don't know about with things they don't know about. So, Neale asks, is it really plausible to attribute aphonic-involving referential intentions to speakers at all? Keeping with the example from before, it is fairly clear that ordinary speakers don't appear to have conscious beliefs or intentions about things like PRO. Linguists didn't discover PRO until very late in the history of natural languages

where PRO is allegedly part of some sentences, so speakers' access to expressions like PRO is very different from their access to expressions like 'eejit.' The logical conclusion, then, seems to be that positing aphonic-involving intentions is also incompatible with Schiffer's assumption that normal speakers have privileged access to what they mean, intend, and believe in uttering something.

Obviously, then, the degree to which these problems are worrying should match the degree to which the transparency assumption is in fact justified. More precisely, there are at least three hidden assumptions at work here:

A1. Speakers have transparent, privileged access to what they consciously intend, mean, and believe in uttering something.

A2. What speakers consciously intend, mean, and believe in uttering something is identical to what they really intend, mean, and believe in uttering something.

A3. Interpretation and inference are not necessary in understanding what one oneself intends, means, and believes in uttering something, but they are necessary in understanding what others intend, mean, and believe in uttering something.

Intention-based semantics, as promulgated by Neale and Schiffer, is thoroughly wedded to A1–A3. This is unfortunate, not only because it will appear to critics that there is no such thing as intentionalism without broadly Cartesian views of the mental, but also because there are good reasons to commit transparency to the flames. Or so I will argue, along lines essentially similar to Neale's "Tacit States Reply" to the meaning-intention problem (2016, §14.12).

4.1. *Aphonic-Involving Intentions*

Let's start with the alleged problem with aphonic-involving intentions before approaching the broader question of meaning-intentions. As stated, the problem is that speakers would be supposed to perform acts of meaning, saying, referring, etc., without knowing anything at all about the means with which they do so. Clearly, PRO and other aphonic expressions will fit the bill; normal speakers don't appear to have any conscious knowledge that such entities exist. On the other hand, speakers appear to know that words exist and they appear to know that words and sentences are means by which they express and communicate their thoughts and beliefs to others.

That may indeed sound reasonable, but only because it is part of folk linguistics and general common sense opinion. If any account of the nature of words along the lines of §3.2 above is correct, normal speakers have no idea what a word is, because they have no idea what a lexeme, root, or morphosyntactic unit is. But this objection is too quick. Surely, no one would suggest that naïve speakers know the true metaphysics of words, what they do know is that there is something in the

world, namely utterances of words and sentences, with which people communicate. And nothing similar can be said about their knowledge of utterances of aphonic expressions.

The deeper problem with this argument is that it draws an illicit distinction between words on the one hand and aphonics on the other. For if we are to accept aphonics at all, they will belong in the category of words; they are merely words without phonetic or phonological properties. So, if we really want to take folk linguistics seriously, naïve speakers will turn out to know about aphonics in virtue of their admittedly superficial knowledge of words in general. Neale, it seems, would have to draw some principled distinction between phonic expressions and aphonic expressions. This would allow him to hold that normal speakers have only encountered utterances of the phonic bit of the lexicon and, so, they only have knowledge of words as phonic entities. Aphonics are completely beyond their ken.

Apart from the problems already mentioned, especially the point that normal speakers may have encountered aphonics in speech by witnessing acts of omitting to say something explicitly, drawing this distinction is not as easy as it seems. To see this, consider some of the most basic theoretical commitments of linguists who pursue an intentionalist theory of phonological competence. Bromberger and Halle (2000), for example, argue that phonological descriptions or derivations of dated utterances should not be understood merely as phonetic transcriptions of a speech event—i.e. symbols encoding articulatory movements—but as standing for a sequence of intentions which give rise to those movements. Simplifying dramatically, the IPA transcription [ɹʷɛd], occurring in a phonological derivation of an utterance, doesn't merely record the phonetic segmentation of an event of uttering 'red' into the three stages of (i) labialized postalveolar approximant, (ii) open-mid front unrounded vowel and, finally, (iii) voiced alveolar stop. It represents the phonetic intention that called for this complex sequence of articulatory movements and positionings. This kind of intention is grounded in linguistic competence, since speakers can very well make the requisite sounds and movements encoded by [ɹʷɛd] without having the intention to utter a word of English, for example if they don't know the language but just happened to produce the sound in the manner required. Phonetic intentions are intentions, then, to produce speech sounds of specific languages.

But what, more specifically, are the objects of phonetic intentions? According to Bromberger and Halle (2000: 26–27), speakers must, at least, have intentions to produce morphemes. A morpheme is either a stem or an affix of a word; 'an-arch-ic', for example, has two affixes, 'an' (prefix), and 'ic' (suffix), and one stem, 'arch'. A single stem, on this kind of theory, can be pronounced differently in different linguistic environments. The stem 'sell' is sometimes pronounced /sold/ and sometimes /sells/, depending on tense and Case agreement. So, whenever I intend

to produce a phonic word I retrieve information about each morpheme from memory, and utter the resulting combination or transformation. Now, we have already attributed intentions to speakers which will be completely unrecognizable to them. Normal speakers usually do not, and need not, know anything at all about the morphemic structure of the words they use. Neither do they need to know that morphemes exist; and they normally don't know. Nevertheless, very basic commitments in (some of) phonological theory involve the attribution of intentions to pronounce morphemes to ordinary speakers. It seems, then, that the assumption of transparency would require a wholesale rejection of these ideas, or very radical revision of foundational assumptions. Neither option is appealing, since dropping transparency seems the easier thing to do (see next section).

Before moving on, it should be noted that adding just one layer of complexity into this analysis will make phonetic intentions to utter phonics appear just as problematic as phonetic intentions to utter aphonics. Plural and past-tense affixes in English have radically different phonological features in different environments. Consider the examples from Bromberger and Halle (2000: 27):

Plural morpheme: cat/s, child/ren, kibbutz/im, alumn/i, stigma/ta, geese, moose
Past-tense morpheme: bake/d, playe/d, dream/t, sol/d, sang, hit

To explain these irregular affixes, they propose a category of abstract morphemes symbolized with the letter Q. Q has no direct phonetic interpretation but, as I understand the idea, it encodes information about how the morpheme is pronounced (it is an 'identifying index'). Bromberger and Halle take care to note that 'Q' is part of the notation of the theory, and not a symbol that really occurs 'in the mind' of the speaker. But this is at best an admission that we simply lack knowledge in this area, since they expect more work in linguistics will eventually reveal mental structures corresponding to representations in the notation of the theory (2000: 26n10). The bottom line, however, is that speakers of English do utter plural and past-tense morphemes, and they must then, in some sense to be explained, intend to utter Q. But it appears that Q either doesn't encode any phonological information, because the phonetics of Q are so radically dissimilar on different occasions of utterance, or the information is almost impossible to specify, even theoretically. Further, it is part of one fairly influential theory of phonological competence that when speakers intend to utter some phonics, they must intend to utter abstract morphemes like Q. Thus, I conclude, it has not been shown that intentions to utter aphonics have to be more problematic than intentions to utter phonics.

4.2. *Access to Propositional Attitudes*

The argument, so far, may seem to amount to no more than simple buck-passing. Surely, one would like to say, there is a world of differ-

ence between conscious, personal-level intentions like the intention to mean that *p* by uttering X and any subpersonal intention or other mental state involved in knowing the grammar of a language. And so, the thought continues, we haven't solved any problem by attributing, say, phonetic intentions to speakers and hearers. Indeed, there is much reason to think that this is merely a *façon de parler* awaiting elimination when science progresses (Collins 2007, 2008). True, this is a popular and plausible way of thinking about these issues but, I want to argue, the problems involved in personal-level intention attribution are much more pressing and consequential than many philosophers of language have hitherto allowed for. Possibly, speakers are built to consciously represent themselves, to themselves, as having certain intentions and beliefs, without this being a good indicator that those are the intentions or beliefs that they actually have. If so, talking in terms of 'conscious intentions' is also just a manner of speaking awaiting elimination.

Start with the nutshell description of Gricean intention-based semantics, or intentionalism for short, coming from Neale, Schiffer and others. According to intentionalism, what the speaker says and means by making an utterance on an occasion is metaphysically determined by certain specific audience-directed intentions. What is strictly *said* by the speaker may need to conform, in some sense, to the linguistic meaning of the sentence uttered, but what is otherwise meant— e.g. conversational implicatures—can roam freely from any such constraint. So, for instance, if I utter,

(8) Meet me at the bank,

the only facts that can determine whether I meant a river bank or a financial institution are facts about my communicative intention at the time of utterance. The context could be such as to make it appear to my audience that I meant the river bank—if I'm holding a fishing rod, for example— even if I really intend to refer to a financial institution, and so they might very well misunderstand me. And in many such cases the responsibility for the misunderstanding falls squarely on the speaker's shoulders; they failed to take the full context into consideration before speaking. But it is still the intention that determines which interpretation is the correct one. Neale (2005: 179–180) describes this in terms of an epistemic asymmetry between the speaker and hearer. Speakers *know* what they mean and the hearer's job is to *work it out*. The speaker normally doesn't need to work this out, they simply know what they mean without interpretation or inference.

Already, I believe, it is important to pry this apart. Even if it is conceded that speakers normally know what they mean, or some part of it, this must be flagged as a thesis in the epistemology of interpretation. According to intentionalism, what I said in uttering (8) on an occasion is determined by my communicative intention, it is not determined— except in the epistemological sense of that word—by what I believe I intended to say. Neither is it determined by what I believe my intention

is while uttering (8). Sure, I could find out what my intention was in uttering (8) by forming a belief about the intention. The difference between the hearer and myself here is, at least, that I often have access to more data—my own mental imagery at the time for example—and may often form my belief without waiting to hear the words I utter. Perhaps this higher-order belief is formed automatically and unconsciously, but it can surely be mistaken like any other belief. Note, however, that more data does not necessarily result in more reliable judgment, as it might just overwhelm one's cognitive system and lead to an increased number of errors. Sometimes, cognitive processes are more reliable if they use only a limited collection of evidence (Gigerenzer et al. 1999; Carruthers 2011: 24). Further, as Peter Carruthers (2011, Ch. 2.5; also pp. 12, 22, 70) has argued, Cartesian transparency derives no support from phenomenological observation, contrary to widespread opinion. Beliefs about what others mean in uttering something are formed just as automatically and sub-consciously as beliefs about what we ourselves mean, intend, or believe. And we have often formed those beliefs as hearers, automatically and predictively, before the speaker puts a stop to the sentence or even before they say anything. And this is predicted by so-called 'forward models' of human cognition (e.g., Pickering and Clark 2014). Normally, people simply find themselves with beliefs about what speakers meant, with no insight into how exactly the belief was formed. We seem also simply to find ourselves with such beliefs about our own propositional attitudes.

4.2.1. *Arguments from Modularity*

As Carruthers (2011) argues at length, there are good empirical and theoretical reasons to think that the human mind—or a specific mind-reading module in the mind—automatically adheres to cognitive procedures which assume the mind is transparent to itself. If so, it is just a near-universal assumption of humans that if one thinks one is in mental state M then one must actually be in mental state M and, also, that if one thinks one is not in mental state M then one must really not be in that mental state (ibid., p. 12). This cognitive procedure is compatible with the suggestion that, in fact, people routinely confabulate and misinterpret their own mental states. And experimental findings indicate that when such confabulation occurs, people don't have subjective access to the information that their self-attribution of a mental state was a complete fabrication. Carruthers takes, as an example, research on commissurotomy patients, where different stimuli are presented to the two hemispheres at the same time.

> The patient fixated his eyes on a point straight ahead, while two cards were flashed up, one positioned to the left of fixation (which would be available only to the right hemisphere) and on to the right of fixation (which would be available only to the left hemisphere). When the instruction, "Walk!" was flashed to the right brain, the subject got up and began to walk out of the

testing van. (The right hemisphere of this subject was capable of some limited understanding of words, but, had no production abilities.) When asked where he was going, he (the left brain, which controlled speech-production as well as housing a mindreading system) replied, "I'm going to get a Coke from the house." This attribution of a current intention to himself was plainly confabulated, since the actual reason for initiating the action was accessible only to the right hemisphere. Yet it was delivered with all of the confidence and seeming introspective obviousness as normal. (2011: 39–40)

Even if patients are reminded, and made fully aware, that the surgery can have effects on their access to their own mental states, they still insist that they know for sure what they really intend to do. As Carruthers emphasizes, this does not support total skepticism about self-knowledge, but it does show that confabulated mental states will appear just as transparently accessible to us as their authentic counterparts.

Now, let's look again at the transparency assumptions A1–A3, starting with the final one.

A3. Interpretation and inference are not necessary in understanding what one oneself intends, means, and believes in uttering something, but they are necessary in understanding what others intend, mean, and believe in uttering something.

As already noted, this assumption gets no support from intuition or first-person phenomenology, since the cognitive process by which we form beliefs about the intentions of others is just as immediate and automatic as that by which we form such beliefs about ourselves. But more substantively, as Carruthers (2011) argues, if one likes the idea that the mind houses a mindreading module specifically geared towards the task of attributing mental states to intentional agents, some very serious empirical arguments can be mustered against A3. I will only give the flavor of these arguments here, and go on to focus more directly on assumptions A1 and A2.

First, assume that there is a mental module for mindreading. Secondly, we can then ask: What is our best theory of the nature and evolution of this module? This is where Carruthers would introduce his interpretive sensory-access (ISA) account of mindreading, according to which the module only has sensory access to its domain, and this access is always—with two notable exceptions, namely aspects of perceptual and emotion-like states (2011, Ch. 4, 5)—interpretive rather than transparent. And so, the ISA theory predicts that when one attributes propositional attitudes traditionally so-called—intention, belief, judgment, desire, etc.—one must engage in interpretation in both self-attributions and other-attributions. This prediction derives some support from the observation that a mindreading module is most likely to have evolved in response to strong social pressures on individuals to acquire capacities to predict and explain the complex and varied behavior of other individuals. If this is in fact the most plausible evolutionary story

we can tell, the simplest hypothesis, according to Carruthers, "…is that self-knowledge is achieved by turning one's mindreading capacities on oneself" (2011: 65). And this, in turn, would suggest that our access to our own mental states is, in essence, the same as our access to the mental states of others, namely by means of inferences based on various sensory cues. In our own case, the data pool will usually be very different, however, involving many private mental episodes and access to personal memory.[8]

4.2.2. *Arguments from Dual System Psychology*

Whatever one thinks about modules for mindreading, there is a strong case to be made against Cartesian transparency on the basis of dual system theories in psychology, or 'fragmentational' theories of mind more generally. By now it is a fairly standard view, but certainly not universal, in the cognitive sciences that the human mind has a fast, intuitive, automatic, and nonconscious processing component and a more effortful, reflective, and conscious component which is subject to voluntary control (cf. Evans and Frankish 2009).[9] Call the first 'System 1' and the second 'System 2'. There is much controversy about how exactly the two Systems relate to one another and how they ought to be defined. System 1, or parts of it, is evolutionarily ancient and is shared with some non-human animals. System 2 is thought to be more distinctively human or, at least, more developed in humans than in other animals. Either the Systems are distinct capacities with different evolutionary histories and physical realizations or they are different ways of utilizing the same cognitive resources, with System 2 operations partly realized in System 1 processes (Carruthers 2009; 2011: 98–101).

Either way, however, if mental partition of this sort is granted, we can inquire into the characteristics of content-bearing mental states as they occur in System 1 or System 2 processing respectively. Philosophers have recently, for example, been very interested in cases where people sincerely profess to deeply held attitudes—e.g. egalitarianism, anti-racism—while unreflective, 'System 1-based' behavior and cognition seems to manifest diametrically opposed attitudes (Gendler 2008; Schwitzgebel 2010). The possibility of such cases should not be taken as proof that there are two mental Systems or two fundamentally different species of propositional attitude (Gigerenzer and Regier 1996;

---

[8] Of course it should be noted that this story is contested, most obviously by simulation theorists like Goldman (2006) who argue that self-directed metacognitive abilities are evolutionarily prior to other-directed mindreading abilities.

[9] If such a heavy-duty psychological theory, much of which is hotly contested, is not allowed as an assumption, we could make do with a fragmentational or partitive theory of the mind (as in Lewis 1982; Davidson 1982; Egan 2008; Mandelbaum 2015). All we really need is the idea that propositional attitudes can be causally isolated from one another within a single mind, making it possible for a single mind to harbor inconsistent beliefs or intentions at any given point in time. Many theorists find the System 1/System 2 distinction relatively intuitive, so I use it here.

Mandelbaum 2015), but they are helpful in understanding the status and interaction of different kinds of content-bearing states of mind. To fix ideas, call contentful mental states occurring in System 1 processes 'A-attitudes' or 'A-states' and those occurring in System 2 processes 'B-attitudes' or 'B-states'. A-states are ancient, automatic, and (perhaps) associative while B-states are bookkeepers who, normally lagging behind, strive to be boss over A-states. Tamar Gendler (2008) calls A-state beliefs 'aliefs' and B-state beliefs 'beliefs' but I prefer my own terminology for sake of generality; now we get to talk about A-intentions vs B-intentions, and so on.

Start with beliefs. Suppose an individual S has the B-belief that *not p* and seems also to know full well that S has that very belief. Of course, S need not realize that the belief is a B-belief since S may not make the A/B-distinction. It's possible, then, that S also has the A-belief that p, directly contradicting the content of S's B-belief. For example, S could be an implicit or unconscious racist, A-believing that other races are inferior, while B-believing that they are not. S's B-belief will consist in things like S consciously and intentionally professing to egalitarian and anti-racist attitudes. S may even rehearse, in inner speech, the conscious B-belief that other races are not inferior. But, surely, this is compatible with the opposing A-belief being manifested in reasoning, action, and automatic reactions to relevant situations. Finally, there is no longer a clear sense in which S knows that S believes that other races are not inferior. For, one might think, genuine belief should require at least the presence of A-belief. We might say instead, in this kind of case, that S merely knows that S professes to believe that *not p* while really believing that p. Often, the real belief will turn out to be 'nonconscious,' but this doesn't seem necessary (Frankish 2016; Hunter 2011).

As Keith Frankish (2004) and Carruthers (2011) have argued, using different terminology, B-beliefs are much more like *commitments* than truthful *reports* of contentful mental states. When I say to myself, or to others, that I believe p, I commit myself to this belief. More specifically, if I happen to take such commitments seriously, I measure myself according to the standard of believing that p: I aim to make my behavior expressive of the belief, experience disappointment when this fails, and exhort myself to stand by my word. Importantly, however, even if my commitment results in everything appearing as if I really believe p, commitment is not identical to belief. Unless, perhaps, one is a full-blown instrumentalist or anti-realist about content-bearing mental states. Let's commit to ignoring such views for the time being. To see the point, one just needs to note the relativity and variability of commitment-attitudes across different individuals and across different times for the same individual. Many people routinely commit to things without having any apparent control over relevant behavior or reasoning. So, it seems, the causal profile of commitment-attitudes depends entirely on which other higher-order attitudes are held by the person in question; in the case of belief it may depend on whether they believe

they actually have made the commitment, whether they want, generally, to make good on their commitments, and so on.

This immediately contrasts with the causal profile of belief, where the question of whether one really has a belief does not depend on attitudes of this kind. And, Carruthers (2011: 102–107) argues, the same thought applies to decisions, judgments, wonderings, and supposings. An actual decision to do something should "settle the matter" by itself, while saying to oneself that one will do it—committing to it—only does so in concert with the right beliefs and desires. Surely, committing to the truth of p may result, in due course, in the A-belief that p. And, generally, self-attribution of the belief that p may be self-fulfilling in that it gives rise to various pressures on one to behave *as if* the attribution is true. But committing and believing are still importantly different. Further, one's commitment-attitudes are seriously vulnerable to systematic and immediate confabulation as well as self-directed propaganda and deception, for example in the service of perpetuating a certain self-image (Wilson 2002).[10] I may commit to p being true because of wishful thinking, B-believing sincerely that I A-believe that p (assuming I have the concept of A-belief), while not really A-believing anything of the sort.

Now, how could this apply to cases like appearing to know what one means in saying something on a particular occasion? Surely this is different from having automatic or non-introspectible attitudes; when one means that p by uttering X one consciously intends to mean that p and has fairly reliable knowledge that this is so, right? Remember, however, that the point is not to show that speakers never know what they mean, only that their access to what they mean is interpretive, inferential and prone to error, just like their access to what others mean. And what I have tried to show is that even if speakers know what they B-intend, it doesn't follow that they know what they A-intend. Further, it is quite likely that the contents of A-intentions differ radically and systematically from the contents of B-intentions, even when both are occurrent attitudes a speaker has in making an utterance. So, the idea goes, speakers may A-intend the proposition that p while actually B-intending the proposition that q. What counts as successful interpretation, on this kind of view, is not completely obvious. But one possibility is that recognizing the A-intention is sufficient by itself, while recognizing the B-intention is not, because it only gives the hearer knowledge of what the speaker consciously believes about the content of the intention. But surely, this is often a good indicator of belief-contents, or some

---

[10] In Kent Bach's (1981) terminology, believing that p is different from thinking that p. So, if I believe that p but want to deceive myself into not believing that p I can fill my mind with thoughts to the effect that p is false, whenever I have occasion to consider my belief. This does not necessarily change my belief, but it may result in self-deception, that is, my conscious thoughts and imaginings will only ever suggest that I believe that p is false, when I really believe it is true. Commitment is more like merely thinking to oneself than really believing.

parts of such contents, and will normally be good enough for purposes of everyday communication. But this is all very abstract. Let's consider two kinds of cases where there is, arguably, an actual mismatch between communicative A-intentions and communicative B-intentions.

Hypnosis. As Carruthers (2011: 342–343; citing Wegner 2002) reports, subjects who are given instructions while hypnotized will invent new intentions when asked to explain why the do what they do after waking from the hypnotic trance. It is likely, then, that subjects form intentions and make decisions to act while under hypnosis, and those intentions remain active when they regain consciousness. The intention, say, to open the book on the table when you see it, may have been implanted while hypnotized but, when asked why you opened the book, you will automatically form some confabulated belief; that you've always wanted to read Anthony Huxley's *Illustrated History of Gardening*, for instance.

To control for merely pragmatic reasons for subjects to report some reason or other—when they don't really know why they are performing the post-hypnotic action—Carruthers suggests that even ambiguous actions would lead to error prone self-interpretation. So, out of two possible interpretations, the subject would self-ascribe the action which is more plausible in the context, without detecting any mismatch between A/B-intentions. He proposes an experiment where the bodily movements of the subject will be ambiguous between waving goodbye to someone and waving away a bug. But we can also go back to our 'bank'-example above. Suppose we have a hypnotized subject, Beatrix, who is already fairly likely to take money to the bank, and likely to go fishing on the river bank. Suppose, then, that the hypnotist instructs her as follows: "You're holding your daughter's money box to make a deposit into her savings account. When you meet Abigail tell her you're going to the bank, so she can join you there." Let's also assume that this is sufficient to implant, in Beatrix, the A-intention to tell Abigail to come along to the *financial* bank and that this very intention remains active after she emerges from hypnosis. When she is then placed in a situation where the other sense of 'bank' is much likelier to be at play, the current prediction is that she would form a B-intention to tell Abigail that she's going to the *river* bank. Assume, for example, that Abigail and Beatrix are much more likely to go fishing than to go to a financial institution and that, when they meet, they're close to the river bank and Beatrix is holding her fishing rod (as well as her daughter's money box).

A competing description of this case is, surely, that the conscious A-intention to refer to a river bank always supplants the alleged B-intention. But I see this as no more than banging one's Cartesian head against the wall. That is to say, if intuitions of first-person transparency are not allowed to carry weight, both descriptions are at least prima facie plausible, and the issue should be decided by more general theoretical considerations. So, it seems possible that both Beatrix and

Abigail misrepresented and misunderstood the former's (A-)intention when she uttered 'bank'.

Implicature and illocution. According to intentionalism, conversational implicature is determined by the speaker's communicative intention on the occasion of utterance. The question of whether or not I conversationally implicate the proposition in (10) by uttering (9) in a particular context is, thus, answered by finding out whether I actually intended to be understood as talking about an open gas station or a closed one (see Grice 1989: 32).

(9) There is a gas station around the corner
(10) that the gas station is open

Implicature breeds plausible deniability. If it turns out that the gas station is closed and, having been reprimanded for misleadingly implying that it was open, I am free to insist that no such thing as (10) was intended. Importantly, however, I am either telling the truth or not. I may have actually intended to implicate (10), succeeded, and then lied about my intention when I realized (10) wasn't true. Alternatively, I may not have intended to implicate that (10) by uttering (9); that's why my later denial can be plausible, for it could, as far as the hearer knows, be true.

But plausible deniability, in turn, breeds credulous self-deception, or so I argue. If it is easy to deny, when challenged, that something was part of one's intention, it is also easy to believe that it actually wasn't, even when this is a form of self-deception. A single propositional attitude only relates to behavior in concert with other attitudes. So, I will only take the umbrella when I believe it's raining if, among other things, I also don't want to get wet. But suppose I don't really believe it's raining and I erroneously self-ascribe the belief that it's raining. As Carruthers (2011: 94) points out, I can still explain why I didn't take my umbrella, while preserving the basic belief—consisting in a near-universally shared cognitive procedure—that my mind is introspectively transparent to itself. I can simply say, to myself, that I really wanted to get wet or that I briefly forgot about the rain while leaving the house. Similarly, if people are put in a situation where it is better—for their self-image or because of social pressure, say—to form the false belief that they didn't intend to imply something or other, they will probably tend to form exactly that belief, immediately and unconsciously.

Consider, by way of example, an argument between a conservative and a liberal about immigration policy. In conversation, Conn has been advocating tight restrictions on the free movement of labor, while Libby wants to abolish national borders. Without articulating explicitly how the statement relates to the more general issue, Conn says,

(11) But the Polish are hard-working.

In this context, Libby may think that Conn is, by uttering (11), implying or presupposing something like: (i) other groups of immigrants are not hard-working, or: (ii) groups of immigrants should only be allowed

to enter the country if they have certain character traits, hard-working being one of them, or even: (iii) non-whites are not hard-working (this would require the conversational salience of non-white immigrant groups). And it is not hard to conceive that Conn really did intend to imply one of (i)–(iii) but, equally, in order to preserve an anti-racist self-image, he could deceive himself into consciously thinking that no such thing as (iii) was really intended. Conn could think to himself before uttering (11): I won't be implying anything like (iii) because I don't believe (iii). But he might still A-believe something like (iii). So, in such a case, plausible deniability becomes a tool for easy self-deception. Unconscious A-intentions and conscious B-intentions come apart.

Consider also the illocutionary force of an utterance. Rae Langton (2009: 33–34) argues that speakers sometimes perform illocutionary acts they don't actually intend to perform. Taking an example from J.L. Austin (1962), she imagines one man saying to another,

(12) Shoot her,

referring to a woman nearby. According to Langton, the speaker (S) may have intended the utterance merely as advice while the hearer (H) actually takes it as an order. She argues, further, that since illocutionary force ought to be partly defined in terms of conversational uptake, the act of uttering (12) in this context may objectively have the illocutionary force of ordering, rather than advising, regardless of S's intention. This goes against the basic premise of intentionalism, according to which H would simply have misunderstood S's actual intention in taking the utterance as an order, if it was really (intended as) mere advice. And so, on this view, the speech act had the illocutionary force of giving advice.

The point here is not to argue against Langton's description but, rather, to show how the distinction between A-attitudes and B-attitudes affords us with a different perspective on a case like this one. Suppose, for instance, that in some sense S is aware or ought to be aware that the utterance might be understood as an order in the context. S knows, say, that H is somewhat deferential and complaisant to others. Still, S could convince themself, even just momentarily, that uttering (12) in this context is merely giving advice, not issuing a command. So, it seems, S B-intends the utterance of (12) as advice but A-intends it as an order. If this is possible, Langton's description is partly vindicated, even on minimally intentionalist grounds. That is to say, the speech act was really an act of ordering, even if the speaker consciously thought what they were doing was giving a piece of advice. But, on these assumptions, S couldn't have A-intended (12) as mere advice because S A-believes that (12) is more likely to be taken as an order.

Developing this point in the detail it deserves will have to wait for a different occasion. We have, however, established so far that assumptions A1 and A2 of so-called Cartesian transparency are not nearly as safe as intentionalists tend to believe.

A1. Speakers have transparent, privileged access to what they con-
sciously intend, mean, and believe in uttering something.
A2. What speakers consciously intend, mean, and believe in utter-
ing something is identical to what they really intend, mean, and
believe in uttering something.

First, it is doubtful that speakers have transparent access to conscious
intentions (A1) but, even if this is conceded, it is considerably more
doubtful that speakers' conscious intentions are constitutive of their
actual intentions. We tell ourselves all sorts of things, and appear to
act for all sorts of fabricated reasons, without thereby having transpar-
ent knowledge of the contents of our actual propositional attitudes.

Where does this leave us? Well, intention-based semantics, as far
as I can see, still stands as a metaphysical theory of what grounds or
determines the proposition(s) expressed by a speaker in making an ut-
terance on a given occasion. This will be the speaker's communicative
intentions, however exactly those are spelled out in the final theory.
Still, since transparency is a doubtful epistemological thesis, the mean-
ing-intention problem becomes less of a sweeping tool for eliminating
implicit reference than it appeared to be. Surely, the fact that speak-
ers believe that they implicitly refer to locations in saying things like
'It's raining', still constitutes some evidence that this is indeed part of
their communicative intention. And, conversely, their apparent lack
of awareness of intending to refer implicitly to epistemic standards or
modes of presentation gives some reason to think that they don't. This
is simply because people do know something about what they mean,
intend, and believe. But they also know something, and in a similar
way, about what others mean, intend, and believe. And it is quite pos-
sible that some parts of speakers' intended meaning are less noticeable
from a first-person perspective, while being more so from a third-per-
son point of view.

Stephen Schiffer (1992) is clearly concerned with arguing against
philosophers' over-intellectualization of normal speakers of natural
language. He argues that since the nonphilosopher would not even
have access to the *form* of a specification of the property of modes of
presentation postulated by a given theory of belief ascription, it beg-
gars belief to suppose that such a person could intend to refer to the
property (1992: 513). On this view, we are barred from positing parts
into conscious propositional attitudes of which the subject cannot pos-
sibly conceive. I rehearse this here to make three related points. First,
this is only credible as an account of our access to conscious attitudes.
As we have seen, these may only be tenuously related to other men-
tal attitudes. Secondly, intentionalism is in danger of succumbing to
another kind of intellectualism, namely, the intellectualism inherent
in assumptions of transparency. The apparent fact that people have
fairly reliable knowledge of the contents of their mental states is not
sufficient to show that there is no gap between the content and the

knowledge of that content. Otherwise, our similarly reliable perceptual capacities should point in the same direction, but everyone allows that there is illusion and hallucination in that case (Carruthers 2011: 34). Thirdly, intentionlists who accept mental transparency run the risk of trivializing propositional attitude ascription generally. That is to say, if a propositional attitude is composed of two items *p* and *q*—one being, e.g., what is said and the other what is implicated—and subjects are generally worse at consciously detecting *q*-type contents than *p*-type contents, attitude attributions by theorists will be systematically impoverished. Arguably, some people are worse than others in consciously detecting some types of propositional attitudes; especially figurative meanings, emotional attitudes, and conversational implicatures. Take figurative meaning interpretation, for example. The ability to interpret and understand metaphorical and ironical utterances can be severely restricted by low IQ, various brain damages and disorders, schizophrenia, and autistic spectrum disorder (Gibbs and Colston 2012: 286–296).

Further, at least if Carruthers (2011, Ch. 10) is right, the empirical evidence suggests that there is no dissociation between other-directed and self-directed mindreading abilities; that is, whenever subjects are cognitively restricted in their capacity to recognize mental states in others they are also restricted in their capacity to recognize those mental states in themselves. Supposing, then, that there is some property F of propositional attitudes such that normal subjects are not very good at finding out about F, transparency-theorists will tend to believe that F is never really a property of propositional attitudes. But there is good reason, especially given the distinction between A-attitudes and B-attitudes, to think there might very well be F-properties. So, transparency amounts to trivializing the contents of mental states by systematically disallowing any F-type property. Now, I have by no means shown that implicit reference to epistemic standards or to modes of presentation must be F-properties of propositional attitudes. But the argument against that possibility was only, as I understood it, that there couldn't be F-properties or, at least, there couldn't be very complicated and unintuitive F-properties which ordinary speakers couldn't possibly conceive of. It follows from the above, however, that epistemic standards and modes of presentation might very well be implicit parts of propositional attitudes, however complex. So, in that respect, they are in the same boat as PRO, location-variables, and the like.

5. *Conclusion*

I conclude that being a Mad Catter isn't all that bad. Mad Catters believe that it is possible to utter an expression with no phonological properties. They may also believe that only *utterances* of words—not the word-types—have meanings that compose into utterance meanings for the wholes of which they are proper parts. I have argued elsewhere that this is very doubtful (Unnsteinsson 2014), but it is not doubtful be-

cause, as Neale claims, there couldn't be such a thing as an utterance of an aphonic. I also conclude that speakers' lack of conscious awareness that they are using aphonics, and their lack of conscious awareness that they intend to say and/or mean that *p* by uttering something on a given occasion, does not imply that they don't use aphonics or that they don't, on that occasion, say and/or mean that *p*.

References

Austin, J. L. 1962. *How to Do Things with Words*. Oxford: Clarendon.

Bach, K. 1981. "An Analysis of Self-Deception." *Philosophy and Phenomenological Research* 41 (3): 351–370.

Bach, K. 2010. "Refraining, Omitting, and Negative Acts." In T. O'Connor and C. Sandis (eds.). *A Companion to the Philosophy of Action*. Oxford: Blackwell: 50–57.

Brand, M. 1971. "The Language of Not Doing." *American Philosophical Quarterly* 8 (1): 45–53.

Bromberger, S. and Halle, M. 2000. "The Ontology of Phonology (revised)." In N. Burton-Roberts, P. Carr, G. Docherty (eds.). *Phonological Knowledge*. Oxford: Oxford University Press: 19–37.

Carruthers, P. 2009. "An Architecture for Dual Reasoning." In J. Evans and K. Frankish (eds.). *In Two Minds*. Oxford: Oxford University Press: 109–127.

Carruthers, P. 2011. *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.

Carston, R. 2002. *Thoughts and Utterances*. Oxford: Blackwell.

Clarke, R. 2010. "Intentional Omissions." *Noûs* 44 (1): 158–177.

Clarke, R. 2012a. "Absence of Action." *Philosophical Studies* 158 (2): 361–376.

Clarke, R. 2012b. "What is an Omission?" *Philosophical Issues* 22 (1): 127–143.

Clarke, R. 2014. *Omissions: Agency, Metaphysics, and Responsibility*. Oxford: Oxford University Press.

Collins, J. 2007. "Linguistic Competence Without Knowledge of Language." *Philosophy Compass* 2 (6): 880–895.

Collins, J. 2008. "Knowledge of Language Redux." *Croatian Journal of Philosophy* 8 (1): 3–43.

Davidson, D. 1982. "Two Paradoxes of Irrationality." In R. Wollheim and J. Hopkins (eds.). *Philosophical Essays on Freud*. Cambridge: Cambridge University Press: 289–305, repr. in Davidson 2004.

Davidson, D. 2004. *Problems of Rationality*. Oxford: Clarendon.

Devitt, M. 2006. *Ignorance of Language*. Oxford: Oxford University Press.

Dwyer, S. and Pietroski, P. M. 1996. "Believing in Language." *Philosophy of Science* 63 (3): 338–373.

Egan, A. 2008. "Seeing and Believing: Perception, Belief Formation and the Divided Mind." *Philosophical Studies* 140 (1): 47–63.

Evans, J. and Frankish, K. 2009. *In Two Minds*. Oxford: Oxford University Press.

Fischer, J. M. and Ravizza, M. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.

Frankish, K. 2004. *Mind and Supermind*. Cambridge: Cambridge University Press.

Frankish, K. 2016. "Playing Double: Implicit Bias, Dual Levels, and Self-Control." in M. Brownstein and J. Saul (eds.). *Implicit Bias and Philosophy Volume I: Metaphysics and Epistemology*. Oxford: Oxford University Press.

Gendler, T. S. 2008. "Alief and belief." *The Journal of Philosophy* 105 (10): 634–663.

Gibbs Jr, R. W. and Colston, H. L. 2012. *Interpreting figurative meaning*. Cambridge: Cambridge University Press.

Gigerenzer, G. 1999. *Simple Heuristics That Make Us Smart*. Oxford: Oxford University Press.

Gigerenzer, G. and Regier, T. 1996. "How Do We Tell an Association From a Rule?" *Psychological Bulletin* 119 (1): 23–26.

Goldman, A. 2006. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press.

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

Hawthorne, J. and Lepore, E. 2011. "On Words." *Journal of Philosophy* 108 (9): 447–485.

Hinzen, W. and Sheehan, M. 2015. *The Philosophy of Universal Grammar*. Oxford: Oxford University Press.

Hunter, D. 2011. "Alienated Belief." *Dialectica* 65 (2): 221–240.

Kaplan, D. 1990. "Words." *Proceedings of the Aristotelian Society* (Supp. vol.) 64: 95–119.

Kaplan, D. 2011. "Words on Words." *Journal of Philosophy* 108 (9): 504–529.

Langton, R. 2009. *Sexual Solipsism: Philosophical Essays on Pornography and Objectification*. Oxford: Oxford University Press.

Lepore, E. and Stone, M. 2015. *Imagination and Convention*. Oxford: Oxford University Press.

Lewis, D. 1970. "General semantics." *Synthese* 22 (1–2): 18–67.

Lewis, D. 1982. "Logic for Equivocators." *Noûs* 16 (3): 431–441.

Mandelbaum, E. 2016. "Attitude, Inference, Association: On the Propositional Structure of Implicit Bias." *Noûs* 50 (3): 629–658.

Miller, J. Unpublished. "The Metaphysics of Words."

Neale, S. 2016. "Silent reference." In G. Ostertag (ed.). *Meanings and other things:*

Essays in honor of Stephen Schiffer. Oxford: Oxford University Press: 229–344.

Pereplyotchik, D. 2017. *Psychosyntax: The Nature of Grammar and its Place in the Mind*. New York: Springer.

Pickering, M. J., and Clark, A. 2014. "Getting Ahead: Forward Models and Their Place in Cognitive Architecture." *Trends in Cognitive Science* 18 (9): 451–456.

Pickering, M. J., and Ferreira, V. S. 2008. "Structural Priming: A Critical Review." *Psychological Bulletin* 134 (3): 427–459.

Recanati, F. 2010. *Truth-Conditional Semantics*. Oxford: Oxford University Press.

Schiffer, S. 1992. "Belief ascription." *Journal of Philosophy* 89 (10): 499–521.

Schwitzgebel, E. 2010. "Acting Contrary to Our Professed Beliefs or the Gulf Between Occurrent Judgment and Dispositional Belief." *Pacific Philosophical Quarterly* 91 (4): 531–553.

Searle, J. 1978. "Literal meaning." *Erkenntnis* 13 (1): 207–224.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., and Wilson, D., 2010. "Epistemic vigilance." *Mind and Language* 25 (4): 359–393.

Stanley, J. 2007. *Language in Context*. Oxford: Oxford University Press.

Unnsteinsson, E. 2014. "Compositionality and Sandbag Semantics." *Synthese* 119 (14): 3329–3350.

Unnsteinsson, E. 2017. "A Gricean Theory of Malaprops." *Mind and Language* 32 (4): 446–462.

Wegner, D. M. 2002. *The Illusion of Conscious Will*. Cambridge: MIT Press.

Wilson, T. D. 2002. *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge: Harvard University Press.

Is There a Meaning-Intention Problem?

JESSE RAPPAPORT
The Graduate Center, City University of New York, New York, USA

Stephen Schiffer introduced the "meaning-intention problem" as an argument against certain semantic analyses that invoke hidden indexical expressions. According to the argument, such analyses are incompatible with a Gricean view of speaker's meaning, for they require speakers to refer to things about which they are ignorant, such as modes of presentation. Stephen Neale argues that a complementary problem arises due to the fact that speakers may also be ignorant of the very existence of such aphonic expressions. In this paper, I attempt to articulate the assumptions that support the meaning-intention problem. I argue that these assumptions are incompatible with some basic linguistic data. For instance, a speaker could have used a sentence like "The book weighs five pounds" to mean that the book weighs five pounds on Earth, even before anyone knew that weight was a relativized property. The existence of such "extrinsic parameters" undermines the force of the meaning-intention problem. However, since the meaning-intention problem arises naturally from a Gricean view of speaker's meaning and speaker's reference, the failure of the argument raises problems for the Gricean. I argue that the analysis of referring-with offered by Schiffer, and defended by Neale, is defective.

Keywords: Grice, speaker meaning, reference, hidden indexical, meaning-intention problem.

1. *What Is the Meaning-Intention Problem?*

1.1. *The Meaning-Intention Problem: Background*

When confronted with some recalcitrant data, it is common practice in linguistic theorizing to invoke hidden representational structure. Hidden structures abound in syntactic theory, but they are common in semantics as well. If one is unable to generate the intuitively correct truth conditions for a sentence simply by assigning semantic values to the overt parts, one can often resolve this problem by appealing to covert variables or modifiers.

For example, it is notoriously hard to account for truth value judgments regarding belief attributions without invoking such hidden structure:

(1) Alice believes that George Eliot is a man.
(2) Alice believes that Mary Ann Evans is a man.

Let us suppose that Alice has heard the names 'George Eliot' and 'Mary Ann Evans' before, and she thinks they denote two distinct English authors. Then it is plausible to suppose that (1) and (2) might receive different truth values. But a naïve, extensional analysis of proper names would make it difficult to account for this difference. This is in part the motivation for Fregean theories of proper names.

On a Fregean analysis, expressions appearing in intensional contexts do not denote their reference, but instead denote their (customary) *sense*. Since 'George Eliot' and 'Mary Ann Evans' differ in sense, their occurrences in (1) and (2) differ in reference, and hence (1) and (2) may differ in truth value. Thus, although a classical Fregean analysis does not require positing any hidden *structure* in a belief report, it does imply that the proposition meant contains *senses*, entities of which speakers may lack a concept.

A competing account can be found in the writings of neo-Russellian theorists such as Nathan Salmon (1986). These theorists differ from the Fregeans in holding that the semantic value of a proper name is exhausted by its referent, however they explain the apparent failure of substitutivity by positing substantial hidden complexity in belief attributions. In particular, whereas a naïve view might hold that *believes* expresses a two-place relation between a believer and a proposition believed, Salmon proposes that belief attributions in fact express a three-place relation between a believer, a proposition believed, and a propositional *guise* under which the believer accepts the proposition:

(3) *For some guise x*, Alice grasps that George Eliot is a man by means of x and BEL(Alice, that George Eliot is a man, x).

According to this view, the semantic analysis of both (1) and (2) is (3). Thus, both sentences are literally true. However, (2) is highly misleading, and pragmatically conveys something false (e.g, that Alice would assent to 'Mary Ann Evans is a man'). Thus, we are simply judging the sentence according to what is pragmatically conveyed, rather than its semantic content. This view also allows us to account for sentences like,

(4) Alice believes that Mary Ann Evans is not a man,

as follows:

(5) *For some x*, Alice grasps that George Eliot is not a man by means of x and BEL(Alice, that George Eliot is not a man, x).

Alice may accept (BEL) that George Eliot/Mary Ann Evans is a man under some guises, and reject it (or accept its negation) under other guises. This elegant analysis allows one to preserve a simplified, Mil-

lian view of proper names, while accommodating the cases of substitution failure in a fairly systematic way.

However, the neo-Russellian analysis of belief attribution also requires the theorist to posit entities—guises—of which the speaker is ignorant, in the semantic analysis of ordinary belief reports. In addition, it implies that the logical structure of a belief attribution is more complex than an ordinary speaker might recognize. These facts are the basis of a persuasive argument against such theories, first developed by Stephen Schiffer (1992). This argument, known as the *meaning-intention problem*, claims that such analyses cannot be correct, because speakers do not, and cannot, mean what such analyses require them to mean.

Schiffer develops his argument in response to the *hidden-indexical theory* of belief attributions. This view, similar to Salmon's, analyzes belief reports as follows:

(6)   Ralph believes that Fido is a dog.

(7)   (*For some m*) ($\prod$*m & B(Ralph, <Fido, doghood>, m))

On this view, (6) is analyzed as the statement that Ralph believes that Fido is a dog under some mode of presentation (MOP) m, where m satisfies some contextually specified constraint on MOPs $\prod$*.

"The meaning-intention problem" in fact covers a number of related concerns raised by Schiffer for this type of analysis:

(a)   The Awareness Problem—If the hidden-indexical theory of belief reports is true, then speakers lack full, conscious awareness of what they mean, for most speakers have no idea that they are referring to a MOP-property: "Thus, if the hidden-indexical theory is correct, then [the speaker] has no conscious awareness of what she means, or of what she is saying… and this is a prima facie reason to deny that she means what the theory is committed to saying she means" (Schiffer 1992: 514).

(b)   The Cognitive-Resources Problem—If the hidden-indexical theory is true, then what a speaker means may involve entities about which the speaker is totally ignorant (such as MOP-properties).

(c)   The Specificity Problem—If the hidden indexical theory is true, then there must be some *particular* property $\prod$* of MOPs that the speaker is referring to. But it's doubtful that the speaker's intentions serve to pick out any such particular property (the speaker lacks "specifying intentions," we may say).

On the assumption that speakers *do*, generally, have full, conscious awareness of their own speech and thought contents, and that speaker-meaning requires a speaker to have a conceptual grasp of the particular entities and properties that comprise the contents of their speech acts, the meaning-intention problem raises serious doubts about the hidden-indexical theory of belief reports. Similar doubts would arise for any theory that posits, in the semantic analysis of some sentence, entities,

properties, or structure about which the speaker (a) lacks conscious awareness, (b) is totally ignorant, or (c) lacks specifying intentions.

Schiffer (1996) deploys this argument against another prominent contextualist theory that invokes hidden indexicals: contextualist theories of knowledge claims. Such theories are in part an attempt to respond to the fact that a sentence like 'I know that I have hands' might be judged true in an ordinary context, but false in the context of a discussion about Descartes's evil demon. Contextualists claim that knowledge claims are tacitly relativized to a *standard of knowledge*, which affects the kind of evidence needed to support a knowledge claim in a given context. Thus, 'I know that I have hands *relative to a low standard of knowledge*' might be true, while 'I know that I have hands *relative to a high standard of knowledge*' might be false.

Schiffer sees two problems with this approach. First, it again implies that speakers are not fully aware of what they mean, think, or say: "But no ordinary person who utters 'I know that *p*', however articulate, would dream of telling you that what he meant and was implicitly stating was that he knew that *p* relative to such-and-such standard" (Schiffer 1996: 326–327). Second, it implies that speakers' conception of their own speech contents may not only be *incomplete*, but may be seriously mistaken. For the contextualist hopes to explain why the following kind of argument *appears* to be sound, even though the conclusion seems incorrect:

(A)   I don't know that I'm not a brain-in-a-vat.
(B)   I know that I have hands only if I know that I'm not a brain in a vat.
(C)   Therefore, I don't know that I have hands.

The response is that (C) strikes us as false because we are prone to interpret it as:

(C*)   Therefore, I don't know that I have hands *relative to a low standard of knowledge*,

even when the context dictates that we ought to interpret it as:

(C**) Therefore, I don't know that I have hands *relative to a high standard of knowledge*.

The problem, according to Schiffer, is that it's highly implausible to suppose that speakers could be so confused about the content of their own speech acts: "It's as though a fluent, sane, and alert speaker, who knows where she is, were actually to assert the proposition that it's raining in London, when she mistakenly thinks she's asserting the proposition that it's raining in Oxford" (Schiffer 1996: 326).

1.2. *Assumptions Behind the Meaning-Intention Problem*

As we have seen, the meaning-intention problem attempts to cast doubt on contextualist semantic theories by arguing that we can't take

speakers to mean what such theories require them to mean. However, the argument is deployed with different emphasis in different places. Thus, it is worthwhile to clearly articulate the assumptions upon which such an argument rests:

(Awareness) If a speaker, S, performs an utterance, and means thereby that p, and x is a constituent of the proposition p, then S must be consciously aware that x is part of what she means.

(Cognitive Resources) If a speaker, S, performs an utterance, and means thereby that p, and x is a constituent of the proposition p, then S must have a concept of x, whether consciously accessible or not.

(Specificity) If a speaker, S, performs an utterance, and means thereby that p, and x is a constituent of the proposition p, then S must have specifying intentions that uniquely determine reference to x, rather than any other nearby candidates that might appear to serve just as well in the context.

Here, I adopt a structured-proposition view of propositional content for ease of exposition. Saying that x is a constituent of the proposition p is simply an attempt to capture the intuitive idea that p is, in some sense, "about" x. If a speaker utters 'It's raining' and means thereby that it's raining in London, then she has said something about London. The truth conditions of her utterance are sensitive to how things are in London.

Although these principles are distinct, they all attempt to use the fact that meaning is an intentional act performed by speakers as a way to constrain what counts as an acceptable semantic analysis. In Section II, I argue that these principles are not sound, and that the meaning-intention problem, as stated, is far too powerful.

1.3. *The Aphonic-Intention Problem*

Stephen Neale (2016) defends Schiffer's meaning-intention problem, and claims there is an additional problem related to the use of aphonic referring expressions. The problem is based on the Gricean analysis of what is it to refer *with* an expression:

(RW) In uttering x, S referred to o *with* (or *using*) e, relative to its i-th occurrence in x, iff for some audience A and relation R, S intended A to recognize that $R(e, x, i, o)$ and, at least partly on the basis of this, that S referred to o in uttering x.

R, in this definition, is, effectively, the *inference-base* feature (cf. Schiffer 2017) of the expression e, that is, the property that S believes e has (relative to its position i in the sentence x), such that the audience will recognize that S is referring to o partly on the basis of recognizing this feature. Often, this feature is simply the fact that e is conventionally used to refer to o.

This definition of *referring-with* invokes the more basic notion of *speaker-referring*, which Neale follows Schiffer in defining as:

(SR) In f-ing, *S referred* to *o* iff what *S meant* by f-ing is an *o*-dependent proposition (a singular proposition that has *o* as a constituent).

This definition, in turn, analyzes speaker-referring in terms of the more basic notion of *speaker's meaning*, and hence, together these definitions allow us to explain what it is for a speaker to refer *with* an expression in basic Gricean terms.

What's important, for the present discussion, about this Gricean analysis of referring-with is that it entails that if a speaker uses *e* to refer to *o*, then the speaker must have an intention that *has the expression* e *as part of its content*. It's not simply that, in referring to *o*, a speaker must employ the expression *e* in actualizing her intention, the way she must perform an alveolar stop in pronouncing the word *to*, something which most English speakers are capable of, whether or not they have a concept of *alveolar stop*. For in general, it isn't the case that if actualizing some intention requires a subject to utilize some capacity, then she must be aware of possessing that capacity. However, due to the nature of communicative intentions, if a speaker uses some expression *e* to refer to *o*, then she must intend for the audience to recognize *that* e *has some inference-base feature*, and thus, she must be capable of forming intentions whose content involves the expression *e*. This implies that referring with some expression *e* requires the subject to have some level of conceptual grasp of the expression *e* itself.

The problem that Neale raises is that this consequence is extremely dubious for cases of aphonic referring expressions, since most speakers have no idea that such expressions exist. Thus, there is a meaning-intention problem not only with respect to the entities that a speaker refers *to*, but also with respect to the entities that speakers refer *with*: "So an implicit reference theory according to which speakers refer aphonically to mode of presentation types faces a compound problem: the theory has ordinary speakers referring *to things* they don't know about *with things* they don't know about" (Neale 2016: 154).

This variation of the meaning-intention problem appears to rely on the following assumption, which is a fairly straightforward consequence of (RW):

(Syntactic Knowledge) If a speaker *S* refers to some entity *o* with some expression *e*, then *S* must be able to form an intention that has *e* as part of its content, and hence must "know about" *e*, in some sense.

I will argue that this assumption is incompatible with the data.

2. *Extrinsic Parameters*

There are a number of examples that appear to contradict the proposed principles supporting the meaning-intention problem. They seem to show quite clearly that a speaker *S* can express a proposition *p* that is about, or concerns, some entity *e*, even though *S* either lacks a concept of *e*, *contra* (Cognitive Resources), or simply lacks conscious awareness of expressing a proposition about *e*, *contra* (Awareness).

(8) This book weighs five pounds.
(9) The flash and the bang happened at the same time.
(10) It's summer.

Each of these sentences expresses a proposition whose truth value depends on the state of some entity that is not explicitly mentioned in the sentence itself:

(8*) This book weighs five pounds (*on Earth*).
(9*) The flash and the bang happened at the same time (*relative to Earth as a frame of reference*).
(10*) It's summer (*in the Northern hemisphere*).

Most educated speakers know that weight is something that can actually vary depending on what large body of mass you happen to be standing on. And many (though perhaps not most) speakers know that simultaneity is not an absolute relation, but rather that two events are simultaneous only with respect to a reference frame.

What's important for our discussion is that despite the fact that many educated speakers are aware of such relativization, *many competent speakers are not*. Indeed, for large periods of human history, *no competent speaker* grasped that weight is relativized or had any idea what a reference frame was. This did not in any way prevent them from communicating complete thoughts in uttering sentences like (8)–(10). Therefore, *contra* (Cognitive Resources), it is not the case that if a speaker expresses some proposition *p* that contains some entity *o*, then she must have some conceptual grasp of *o*. *A fortiori*, she need not have conscious awareness of such a concept either, *contra* (Awareness).

Considering sentence (10), even if we assume that all, or almost all, competent speakers know that what season it is depends on what hemisphere you are in, it is entirely plausible that a speaker might utter (10) and not be thinking about hemispheres at all, i.e., not have any awareness of saying something that is about a hemisphere. This again contradicts (Awareness). Even for those cases in which speakers have the requisite concepts, they may not be aware of invoking those concepts in performing their utterance.

Neale (2016: 160ff.) raises similar examples, referring to such cases as involving "extrinsic parameters," but fails to draw any substantial conclusions from them. Instead, he focuses on the disanalogies between these cases and MOPs, in order to show that whatever solace the hidden indexical theorist hopes to find in such cases does not help the hidden indexical theory of belief reports. Examples (8)–(10) relate to

> factors *external* to us… about which we may be ignorant but about which we may acquire knowledge and thereby easily refine our linguistic behaviour. Mode of presentation types are not like this at all. They are supposed to be things *under* which beliefs are *had*, and learning about their existence and a great deal of information about their roles in theories of language and mind doesn't even put theorists in a position to articulate the truth conditions of the propositions they actually express on given occasions us-

ing belief sentences if the hidden-indexical theory of belief reports is true. (Neale 2016: 163)

Thus, unlike with MOPs, "once speakers *learn* about time-zones, hemispheres and rest-frames, and learn a few additional words, they can easily describe the parameters relevant to the truth or falsity of what they are saying…"

First of all, this seems doubtful. For instance, I know that whether or not two events are simultaneous depends on a reference frame. However, not having much grasp of relativity theory, my knowledge of reference frames ends there. It is fair to say that I know far more about MOPs than about reference frames.

The main issue, though, is not whether MOPs are easier to grasp than reference frames, or *vice versa*. The problem is rather that examples like (8)–(10) seriously undermine the general assumptions that support the meaning-intention problem. It isn't clear what comfort the *proponent* of the meaning-intention problem is supposed to find in the fact that even though most speakers lack a proper concept of a reference frame, and many would never "dream of telling you that what he meant and was implicitly stating" was something to do with reference frames, there are *others* who do grasp the concept, and perhaps with sufficient training the rest of the population could do so as well.

The fact is that relatively few competent speakers grasp the concept of a reference frame. Some might be able to recall that simultaneity is a relative notion, if pressed, but even this is reserved for an educated segment of the population. Nevertheless, ordinary speakers are perfectly competent with phrases like "at the same time." This, again, shows that the principles adduced to support the meaning-intention problem cannot be sustained. There may be important differences between MOPs and reference frames, but these differences cannot be used to salvage the meaning-intention problem in its current form.

What about Neale's syntactic analogue of the original meaning-intention problem? Is there an aphonic-intention problem?

I claim, once again by *reductio ad absurdum*, that (Syntactic Knowledge) simply cannot be supported in light of the data. Consider the following sentence of Spanish:

(11)   Quiere comer. [*He wants to eat.*]

According to standard assumptions of generative syntax (cf. Haegeman 1994: 68–69ff.), in order to comply with the Extended Projection Principle, it is argued that (11) must contain a phonologically null subject, which is typically expressed as *pro*:

(12)   *pro* quiere comer.

Pro (distinct from *PRO*) is a phonologically null pronoun, which is the subject of the main clause, and whose semantic value is determined by the speaker's intentions. Importantly, although its existence is not supported by every syntactician, *pro* is an established posit in syntax

with independent syntactic support—in other words, it is not something that is just posited by philosophers in order to generate their desired truth conditions.

Precisely the same concerns that Neale raises for aphonic indexicals in 'Silent Reference' would have to apply to *pro* (which is, effectively, an aphonic indexical). Most speakers lack any conceptual grasp of the expression *pro*, and the speaker of (11) might positively deny that she used a tacit referring expression at all. Therefore, if Neale's arguments are sound, we find that we must reject some basic posits of mainstream syntactic theory as well.

Regardless of whether *pro* exists or not, is doubtful that Neale would be sanguine about this consequence. He often takes philosophers to task for their tendency to make non-trivial claims about syntax purely on the basis of philosophical considerations. But that is precisely what we would have to do if we accept (Syntactic Knowledge).

On reflection, we can see that (Syntactic Knowledge) is a rather demanding principle. It effectively implies that linguistic ability and metalinguistic knowledge must proceed "in tandem"—I cannot *use* an expression to refer unless I have conceptual grasp of that expression. Even for overt expressions, however, it is not obvious that we should accept such a principle.

The problem for the Gricean is that (Syntactic Knowledge) is a fairly direct consequence of (RW). But it appears that (Syntactic Knowledge) must be rejected. Therefore, it seems that (RW) must be rejected as well.

3. *Tacit States*

One direct consequence of the preceding discussion is that (Awareness) should be abandoned. This is all for the good, for the picture of meaning that it presupposes is a strongly Cartesian one. Why should we assume that speakers *do* have privileged access to every aspect of the contents of their speech acts? This is certainly not required in order for them to have meaning-intentions, assuming that such intentions may fail to be fully conscious. And why shouldn't this be the case? Certainly, some strong arguments would be needed to establish that intentions must be conscious; or, at least, that meaning-intentions are special in that they must be conscious. But the Gricean ought to be very cautious about the latter claim, for her entire program is based around positing a certain kind of complex intention as the basis for communication, where this intention does not simply reveal itself through introspection.

The question, then, is whether (RW) can be salvaged by interpreting the intentional verbs *intend* and *recognize* in terms of tacit states—tacit intentions and tacit recognition.

First, it is worth noting that (RW), as it stands, is inadequate—it fails to provide either necessary or sufficient conditions for referring-with:

(i) Consider the following sentence:

(13) I saw Alice$_i$ and then she$_i$ disappeared.

Here I have referred to Alice, and there are two expressions *with which* I referred to Alice. The problem is only the first instance can satisfy (RW). Recall (RW):

(RW) In uttering x, S referred to o *with* (or *using*) e, relative to its i-th occurrence in x, iff for some audience A and relation R, S intended A to recognize that $R(e, x, i, o)$ and, at least partly on the basis of this, that S referred to o in uttering x.

(RW) requires that if a speaker S uses e to refer to o, then S must intend her audience A to recognize that S referred to o in uttering the sentence x, and to do so at least partly on the basis of her utterance of e. But the problem is that A will recognize that S referred to Alice simply on the basis of the utterance of 'Alice.' Since "referring to Alice" really means "expressing an Alice-dependent proposition," then A will know that S referred to Alice as soon as 'Alice' is uttered (and S will know this). Therefore, A will recognize (and S will know that A will recognize) that S referred to Alice purely on the basis of the first reference to Alice—and if this provides a sufficient reason for A to believe that S referred to Alice (as indeed it does), then S cannot intend for A to recognize this even *partly* on the basis of the second reference to Alice ('she'). But, intuitively, S referred to Alice with 'she.' Therefore, (RW) does not provide a necessary condition for referring-with.

To resolve this problem, (RW) must therefore be modified along the following lines:

(RW*) In uttering x, S referred to o *with* (or *using*) e, relative to its i-th occurrence in x, iff for some audience A and relation R, S intended A to recognize that $R(e, x, i, o)$ and to recognize that $R(e, x, i, o)$ *provides a reason* to believe that S referred to o in uttering x.

This modification avoids the preceding worry because it does not require that S intend that A's belief that S referred to o be derived *on the basis of* A's recognition that $R(e, x, i, o)$, but simply that such recognition provide the hearer with *a basis* for arriving at such a belief (whether or not it is the basis that is in fact used). However, (RW*) does *not* require S to even be referring to o (or even to intend for A to recognize that she is referring to o)—it only makes the weaker requirement that S intend for A to recognize that S is doing something that *provides a reason* to believe that S referred to o. This seems too weak. Thus, perhaps the following will suffice:

(RW**) In uttering x, S referred to o *with* (or *using*) e, relative to its i-th occurrence in x, iff for some audience A and relation R, S intended A to recognize that $R(e, x, i, o)$ and to recognize that $R(e, x, i, o)$ provides a reason to believe that S referred to o in uttering x, *and to recognize that S referred to o in uttering x.*

(ii) The second problem with (RW) is that it does not only apply to referring expressions. Consider, e.g., a classroom in which there are two teachers, Alice and Bob, and a number of young students, one of whom is named Bob. Alice is talking to Student Bob's parents and says:

(14) Bob will be teaching the class today,

meaning thereby that Teacher Bob would be teaching the class. In uttering (14), Alice relies on the conventional meaning of *Bob* as a name for Bob, and thus, following (RW), refers to Bob with *Bob* insofar as she expects her audience to recognize that she is referring to Bob by uttering *Bob*. However, she can't expect that uttering *Bob* is *sufficient* for her audience to recognize that she is referring to Teacher Bob, and not Student Bob. Rather, it is the context of the sentence as a whole—or, in particular, the verb 'teaching'—that makes it clear which Bob Alice is referring to. Thus, the right-hand side of (RW) appears to be satisfied by the utterance of 'teaching' as well. (RW) says that a speaker S refers to *o* with some expression *e* just in case for some audience *A* and relation *R*, S intended *A* to recognize that *R(e, x, i, o)* and, at least partly on the basis of this, that S referred to *o* in uttering *x*. Instantiating the variables: Alice intended her audience to recognize that 'teaching' bears some relation to Teacher Bob and, at least partly on the basis of this, that S referred to Teacher Bob in uttering *x*. Thus, (RW) implies that Alice referred to Teacher Bob *with* the word 'teaching.' Since, intuitively, Alice did *not* refer to Bob *with* the word 'teaching,' this implies that (RW) does not provide a sufficient condition for referring-with either.

Unfortunately, no obvious solution to this problem presents itself. The simplicity of (RW) lies in the fact that it analyzes referring-with in terms of offering reasons to believe that one is expressing an *o*-dependent proposition. But, *prima facie*, there is no reason why this condition should be satisfiable by referring expressions only, since other information in the sentence might be intended to help convey what the speaker is referring to, as well.

These problems of definition notwithstanding, the question remains whether (RW) (or (RW**)) is acceptable if one reads the intentional verbs in terms of tacit states.

Two issues must be separated. In discussing tacit states, Neale suggests that 'tacit' amounts to 'unconscious,' glossing "The Tacit States Reply" as assuming that "S is not 'consciously aware' that she means a proposition of the form…" (Neale 2016: 163). This appears to be the sense of 'tacit' used by Brian Loar (1976) as well. However, Neale also describes the view as assuming "tacit knowledge" in Chomsky's conception of the term. Chomskyan tacit knowledge is not merely unconscious, but is also functionally isolated—grammatical knowledge is not integrated into the web of belief, i.e., it is not accessible to central reasoning processes. One might assume that these two categories bear some logical relation to each other—for instance, that anything that is in the web of belief is accessible to consciousness—but this is by no means obvious.

Which of these senses is more appropriate for (RW)? (RW) must grant that speakers tacitly know, e.g., that the Spanish sentence *Quiere comer* has an unpronounced nominal expression in subject position. But it is doubtful that all competent Spanish speakers have a concept (even an unconscious one) of the expression *pro*. Rather, this "knowledge" is more akin to the kind of syntactic knowledge that comprises the language faculty—not fully conceptual, and not integrated into central reasoning. Therefore, the defender of (RW) must grant that the tacit knowledge mentioned in the definition is tacit in the *strong*, Chomskyan sense.

However, introducing tacit intentions in this manner presents a radical break from traditional Gricean thinking. For one of the apparent advantages of Grice's theory is that it accounts for the ways in which linguistic production interacts with global reasoning abilities. Speakers intentionally produce effects in the audience that are responsive to background knowledge, features of the context, and other factors that interact with central reasoning. By claiming that referential intentions can be tacit in the Chomskyan sense (functionally isolated), the defender of (RW-T) weakens the connection between speaker's meaning and rationality. If one pursues this approach to the end, one may find that the result is no longer distinctly Gricean.

4. *Specificity*

I have argued that examples like,

(9) The flash and the bang happened at the same time,

show convincingly that the principles underlying the various aspects of "the meaning-intention problem," in particular (Awareness) and (Cognitive Resources) are not tenable.

If we reject (Cognitive Resources), then we must reject (Specificity) as well. The problem raised by (Specificity) is that it seems like there ought to be something about the speaker's state of mind, or communicative intentions, that determines *which* of a number of competing alternatives is the actual semantic value of a hidden indexical. But if speakers can refer to *o* (express an *o*-dependent proposition) without even having a concept of *o*, as I claim is demonstrated by sentences like (9), then it cannot be the speaker's specifying intentions that resolve the (Specificity) problem.

However, it is worth noting that (Specificity) is not essentially a problem about language or communication, but rather is a puzzle about thought contents. For example, consider:

(15) The book is covered with paint.

Suppose a speaker utters (15) to communicate that the addressee's favorite book, *War and Peace*, which was left sitting on the dining room table, is covered with paint. The problem relating to (Specificity) is that *the book* does not serve to uniquely pick out the aforementioned book,

and there does not seem to be any unique candidate for completing the nominal that is the one that the speaker intended to communicate. This is what is known as an *incomplete description*.

This presents a puzzle about the speaker's *thoughts* about the book, as well. In particular, what is the conceptual content of the speaker's thought, *the book is covered with paint*? One might assume that when a speaker tokens such a thought, there is some description or other (over and above *the book*) under which she thinks of the book. But why would she token one description or the other on any given occasion? Is doing so *necessary* in order for her to entertain thoughts about the book? But surely she needn't token a *uniquely* identifying description of the book, for coming up with such a description is no simple task. But then it begins to look as though the descriptive information is not doing any cognitive work, and the thought is already complete as it is. But if that's the case, then why should an utterance like (15) not be complete as it is?

In sum, I am arguing that (a) examples like (9) show that (Specificity) must be rejected if (Cognitive Resources) is rejected, as I claim it must be; and, (b) the issue of incomplete definite descriptions is primarily an issue regarding mental contents, and only derivatively an issue pertaining to communication. Furthermore, brief consideration of the matter suggests that incomplete definite descriptions may very well exist in thought as well, which implies that the assumptions behind (Specificity) are misguided.

5. *Constraints on Semantic Theories*

If these arguments are correct, then the meaning-intention problem is based on implausible assumptions, and cannot be used as a way to refute the hidden indexical theory of belief attribution, or other hidden indexical theories. Speakers can express a proposition that is *o*-dependent even if they entirely lack the concept of *o*. However, there are significant differences between the Extrinsic Parameters Reply as applied to reference frames and as applied to MOPs and other more speculative philosophical entities.

According to Hofweber (1999), the relevant difference is that when it comes to reference frames (with respect to simultaneity) or bodies of mass (with respect to weight), the extrinsic parameter, or unarticulated constituent, is constant for all members of the speech community. Thus while people failed to realize that weight is a relativized notion, there could never be any confusion or disagreement amongst the members of the community regarding which planet was relevant to their weight statements. This allowed discourse about weight to carry on unproblematically, without the body of mass being specified by the speakers' intentions. This is because the relativization did not affect "sameness and difference (or incompatibility) of contents" amongst speakers.

The same facts do not apply to MOPs. MOPs are *not* shared amongst

all members of the community, and there *is* the tendency for confusion when making belief reports.

Hofweber concludes that the problem with the hidden indexical theory of belief reports is that it implies "not only… that speakers have no access (in the strong sense spelled out above) to the content of their utterances, but also no access to sameness, difference and incompatibility of the contents of their utterances" (Hofweber 1999: 102). However, he leaves it open why this fact is problematic.

I suspect that the problem is ill-formulated as once again pertaining to *access*. The reason why speakers must "share the referent" when it comes to genuinely unarticulated constituents (extrinsic parameters) is that testimony would break down otherwise. For two speakers might use the same sentence to express thoughts that concern distinct entities without their being aware of any difference. Thus, if one speaker asserts that sentence to another, the addressee might form a belief that has entirely different truth conditions. Over time, this discrepancy might cause problems and lead to the extinction of the use of the term.

Alternatively, through the course of such discrepancies, speakers may learn to identify in what ways the truth conditions of each speaker's use diverge from each other, and thus may gain a conceptual understanding of the way that the sentence is relativitized. This might be an important step in coming to grasp the true nature of the relation being discussed.

In sum, defenders of hidden indexical theories for belief attributions and knowledge claims cannot appeal to extrinsic parameters to supply the referents in such utterances, since the alleged referents are not constant amongst members of the community, or across contexts. The only alternative is to argue that speakers really do grasp concepts such as MOPs and their communicative intentions do serve to pick out properties of such MOPs in the required way. This would be to abandon the Extrinsic Parameters approach, but this leads back to the same problems initially raised by Schiffer and Neale.

6. *Conclusion*

The meaning-intention problem has been used by theorists in the Gricean tradition to reject certain semantic hypotheses involving hidden indexical expressions. Hidden indexical theories imply that speakers lack full awareness of their own speech and thought contents, and in some cases lack the relevant concepts entirely. This appears incompatible with the Gricean approach to language, according to which speech act content depends on the speaker's communicative intentions. I have attempted to reverse this dialectic by showing that the assumptions that support the meaning-intention problem are incompatible with some basic data. Since these assumptions are natural ones for the Gricean to make, I have attempted to explore how this affects the project of intention-based semantics. I have argued that the definition

of referring-with that is offered by Schiffer and Neale needs to be re-considered. It fails to provide either necessary or sufficient conditions for referring with an expression, and it is incompatible with facts about *pro*-drop languages, such as Spanish.

The meaning-intention problem, as traditionally conceived, thus appears to have little force. Nonetheless, I agree with Hofweber that appeals to extrinsic parameters must be tightly constrained. The clear cases of extrinsic parameters, or unarticulated constituents, that I have appealed to (such as weight and simultaneity) involve parameters that are constant amongst the members of the speech community. MOPs and standards of knowledge do not share this feature, and thus the same considerations do not apply.

References

Haegeman, L. 1994. *Introduction to Government and Binding*: *Second Edition*. Malden: Blackwell.

Hofweber, T. 1999. "Contextualism and the Meaning-Intention Problem." In Korta, K. et al. (eds.). *Cognition, Agency and Rationality: Proceedings of the Fifth International Colloquium on Cognitive Science*. New York: Springer: 93–104.

Loar, B. 1976. "The Semantics of Singular Terms." *Philosophical Studies* 30 (6): 353–377.

Neale, S. 2016. "Silent Reference." In Ostertag, G. (ed.). *Meanings and Other Things: Essays in Honor of Stephen Schiffer*. Oxford: Oxford University Press: 229–342.

Salmon, N. 1986, 1991. *Frege's Puzzle*. Atascadero, California: Ridgeview Publishing Company.

Schiffer, S. 1992. "Belief Ascription." *The Journal of Philosophy* 89 (10): 499–521.

Schiffer, S. 1996. "Contextualist Solutions to Scepticism." *Proceedings of the Aristotelian Society* 96 (1): 317–333.

Schiffer, S. 2017. "Intention and Convention in the Theory of Meaning." In Hale, B., Wright, C., & Miller, A. (eds.). *A Companion to the Philosophy of Language, Second Edition*. Malden: Wiley Blackwell: 49-72.

Value Disagreement and Two Aspects of Meaning

ERICH RAST
New University of Lisbon, Lisbon, Portugal

The problem of value disagreement and contextualist, relativist and metalinguistic attempts of solving it are laid out. Although the metalinguistic account seems to be on the right track, it is argued that it does not sufficiently explain why and how disagreements about the meaning of evaluative terms are based on and can be decided by appeal to existing social practices. As a remedy, it is argued that original suggestions from Putnam's "The Meaning of 'Meaning'" ought to be taken seriously. The resulting dual aspect theory of meaning can explain value disagreement in much the same way as it deals with disagreement about general terms. However, the account goes beyond Putnam's by not just defending a version of social externalism, but also defending the thesis that the truth conditional meaning of many evaluative terms is not fixed by experts either and instead constantly contested as part of a normal function of language.

Keywords: Disagreement, meaning vectors, externalism, metalinguistic negotiation, truth-conditions.

1. *Introduction*

Within the recent debate about relativist semantics of evaluative predicates and the corresponding notion of faultless disagreement attention has shifted towards more general discussion of value disagreement. The problem is how to account for substantive value disagreements without degrading them to merely verbal disputes. For it seems that if two people disagree about what is good in a given situation, for instance, if they associate different criteria with words like 'good' and 'better than' and one of them says 'Capitalism is good' and the other one replies 'No, it is not', then it might appear as if they only disagree about the meaning of 'good' and in the end only argue about words. In reply to this form of relativism, Alexis Burgess, David Plunkett, and

Tim Sundell (henceforth abbreviated BPS) have argued in a series of articles that such disputes are metalinguistic negotiations about the best use of a word in a context but are nevertheless substantial.[1]

In this article I argue that the metalinguistic negotiation account is incomplete. BPS are right that value disagreements can be metalinguistic in the sense of being disagreements about the meaning of evaluative expressions. These meanings are not negotiated, though, nor is the 'best use' of a word negotiated. Rather, value disputes are instances of ordinary meaning disputes about general or abstract terms and corresponding predicates. If arguments from early semantic externalists like Putnam are taken seriously, such disputes are a normal function of language and there is no substantial difference between the way we disagree about utterances containing the word 'electron' and those that involve a use of words like 'good' and 'capitalism'. However, Putnam's version of externalism is not directly applicable to value disputes because of its strong externalist assumptions. Instead, a dual aspect theory of meaning is suggested that only presumes a weak form of semantic externalism for evaluative expressions and remains agnostic about the strong externalism thesis for natural kind terms. According to my suggestion, speakers often disagree about what an expression *really* means, about what I call its *noumenal meaning*, on the basis of a shared but possibly incomplete *core meaning*. Only mastery of the core meaning is required by virtue of linguistic competence.

The remainder of this article is structured as follows. In Section 2 contextualist, relativist, and metalinguistic theories of value disagreement are laid out. After a brief critique of the metalinguistic negotiation view I introduce the two aspects of meaning mentioned above and show in Section 4 how they explain metalinguistic value disputes. The result should be understood as a précis of the approach of BPS and is defended against possible objections in Section 5. A short summary is given in Section 6.

2. *Contextualist and Relativist Disagreement*

In this section some forms of disagreement are laid out on the basis of distinguishing what two discourse participants disagree about in a traditional truth-conditional approach: Is it the meaning of words, their truth-conditions, or various features of the context and/or the circumstances of evaluation of an utterance?

In order to address this question something must be said about truth-conditions first. As young Wittgenstein put it: "To understand a proposition means to know what is the case, if it is true." (Tractatus, 4.024)[2] In modern versions of truth-conditional semantics it has

---

[1] See Sundell (2011); Burgess (2013); Burgess and Plunkett (2013); Plunkett and Sundell (2013, 2014).

[2] Cit. in Wittgenstein (1969).

become customary to additionally provide a way to deal with indexical expressions such as the tenses, and in the philosophy of language two main traditions have become prevalent. In a Lewisian account based on (Lewis 1980) a sentence φ is true or false in a model relative to a context c and an index i. The context provides a way to fix the denotation of indexicals, whereas the index consists of world-time pairs derived from the time and world of the context.[3] The purpose of the index in this type of theory is purely technical: Expressions like 'yesterday' or 'it is possible that' implicitly quantify over it as part of their meaning. For example, 'it is possible that φ' is true in c,i if and only if φ is true in c,i' for an index i' that is the same as i except that the world of i' is among those that are closest to those of i.[4] Importantly, the world and time of an index is derived in some automatic way from the context.

Kaplan's two-dimensional semantics is similar to Lewis's suggestion in many respects.[5] The main difference is that Kaplan's logic is weakly intensional and two-layered, thereby introducing the notion of the semantic content of an utterance:

linguistic meaning   +   context $\Rightarrow$ semantic content
semantic content   +   CEs $\Rightarrow$ extension

Hereby, 'CEs' is an abbreviation for circumstances of evaluation, the way Kaplan prefers to call the index. In order to avoid confusions with more general talk about modal indices by other authors like Cresswell (1990, 1996) this abbreviation will be used from now on. Kaplan's approach allows for distinctions based on semantic content, which are intensions that play the same role as propositions in one-dimensional theories, and because of this increased expressivity it will be used as a basis of the following considerations unless otherwise noted. However, like in Lewis's approach, CE's play a merely technical role as a means to implicitly quantify over world-time pairs while keeping the interpretation of indexicals rigid. As an example, if Bob says 'Yesterday, I was here' in Pasadena on 5/23/2016, this is evaluated in such a way that the semantic content of 'I' is Bob, that of 'here is Pasadena and the utterance as a whole is true if and only if Bob was at Pasadena on 5/22/2016, the day before the the day of the context.

Both theories allow for several distinct types of disagreement, some of which are *metalinguistic* and some of which are *content-based*. Starting with the latter first, consider the dialogues (1ab) and (1ac):

---

[3] See Lewis (1980: 85–88).

[4] In this formulation, Lewis's neighborhood semantics is assumed. In contrast to this, in a normal modal logic the world of i′ must be accessible from the world of i (= the world of c, in this case, because there are no nested modal operators) by a dyadic accessibility relation. These details make for large technical differences but we can ignore them in what follows, since the nature of the respective modalities plays no role in the following discussion.

[5] See Kaplan (1989). He calls the linguistic meaning of an expression its 'character', but we stick with the term meaning.

(1) a. Anna: Capitalism is good.
     b. Bob: No, it isn't.
     c. Bob: No, Democracy is not good.

The idea of direct content-based disagreement is that Anna expresses some semantic content φ that Bob denies directly in the sense that his reply expresses an ordinary truth-functional negation $\neg\varphi$ of the original content. He makes this explicit in (1–c). Attributing this form of disagreement only makes sense when we have reasons to believe that Anna's utterance and Bob's reply are based on the same linguistic meaning and take place within the same context of use, as the example suggests. In a Kaplanian framework the semantic content of Bob's utterance can then be taken as the negation of Anna's to the effect that he directly contradicts her.

Within these theories there are many other forms of disagreement that are not discussed in the literature very often. Two agents may disagree about empirical facts of the context of use, for example about the question who is the speaker or what the time of utterance was. Two agents may also disagree about contextual factors that are given by facts but rather belong to the cognitive context, to a pragmatic theory of interpretation, and are nevertheless often muddled into the context parameter.[6] For example, they may disagree about the place denoted by a use of 'here' in the following utterance:

(2) Anna: It's cold here.

Is it cold in the bed, cold in the hut, cold in Juneau, Alaska, cold in Alaska at this time of the year, cold on this continent as opposed to Australia, cold on Earth as opposed to Venus, cold in the Sagittarius arm of the Milky Way, cold in this part of the universe, and so forth? The indexical can have any of those intended meanings and the disagreement is not semantic in the sense of being disagreement about a referent that is provided by linguistically-mandated rules. Instead, the disagreement is based on different interpretations when the hearer does not take up the speaker's referential intentions. This pragmatic disagreement about the context can occur even if both speakers agree on the place of utterance. Note that this type of disagreement can also occur in the above example (1), namely about the extension of the present tense that is part of the meaning of 'is'. Anna might want to convey that capitalism is good in general (i.e., the generic reading of the present tense), but she may also intend to convey the assertion that capitalism is good for us now, as opposed to capitalism in Ancient Greece or in the near future.

Two agents may also disagree about the question whether a certain expression is context-sensitive and about which features of the context

---

[6] The term cognitive context is borrowed from Penco (1999) who introduces other useful distinctions. I have laid out my own view about this type of context at several occasions, see e.g. Rast (2009, 2014).

are relevant for determining its semantic content. For example, Anna and Bob may disagree about the question whether the token of 'here' in (2) denotes the hut, because Anna intended the token to do so, and Bob is just wrong about taking it to stand for Alaska, or whether the semantic content of (2) is underspecified, and Bob can therefore claim that his interpretation has some legitimacy, since Anna ignored the fact that they were just talking about the climate in Alaska when she made the utterance. Whatever stance one takes about their dispute, it is hard to deny that their disagreement is metalinguistic. It is also not hard to imagine circumstances under which (1) might constitute a similar dispute *if* 'good' is context-sensitive as many moral philosophers would claim. For example, Anna might argue that she meant 'Capitalism is good for me' and this is the correct interpretation given that it was her intention to convey this message, whereas Bob might argue that 'good' principally has no genuine relational reading, as Moore (1903) famously claimed.

Relativist positions have been developed as alternatives to the classical contextualism outlined above, though within the same truth-conditional setting, and defended for epistemic modals, future tenses, knowledge attributions, aesthetic predicates, predicates of personal taste, and evaluative language in general.[7] The key difference between semantic contextualism and relativism is that in the former any additional parameters of the CEs are derived from corresponding features of the context. MacFarlane (2012) calls a position that puts nontraditional factors into the context and, by some linguistically mandated derivation, into the CEs at later stages of evaluation *nonindexical contextualism*. Consider the following utterance, for example:

(3) Bob: Roller coasters are fun.

If the linguistic meaning of 'fun' is taken to be context-sensitive to an experiencer and this experiencer is determined in the context of use such that the semantic content ends up with a specification of this experiencer, then this is a nonindexical contextualist semantics for 'fun'. The representation is semantic, because the additional factor is mandated by linguistic meaning and fully specified within a context. It is nonindexical, because it need not be claimed that the additional factor is determined by the context of use in the narrow sense *only*. The experiencer need not be the speaker but could rather be determined on the basis of the speaker's intention (e.g. in the reading 'fun for us') or by the conversational context.[8] The representation is contextualist, as long as the experiencer is derived or otherwise determined from the context of use. Hence, in a two-layered model the semantic content will change whenever the respective feature of the context changes.

[7] See for instance Kölbel (2002), Lasersohn (2005), MacFarlane (2005, 2014), Egan (2014).

[8] While there may be concerns about this way of 'semanticizing' possibly pragmatic interpretation processes, these need not worry us here. It is one possible position that may give rise to a particular kind of contextual disagreement.

In contrast to this, according to *assessment-relativism* contexts and CEs are decoupled from each other, and an additional assessor or relevant features of an assessor such as evaluative standards are taken to be a constitutive part of the CEs. This means in the two-layered model that one and the same utterance in a given context of use expresses the same semantic content, but that the truth or falsity of this content may vary from assessor to assessor. So if Anna disagrees with Bob about (3) by uttering 'No, Roller coasters aren't fun', this may be taken as a case of faultless disagreement. (3) in the context c in which Bob is the speaker expresses some semantic content p that leaves the assessor of 'fun' unresolved, i.e., the proposition that roller coasters are fun. This content p then comes out true or false in a model relative to the respective features of the assessor of the CEs. In the example, if the assessor is Anna, then it will be false, and if the assessor is Bob, then it will be true. Since the assessor is not derived from the context of use, Bob's assessment has no priority over Anna's. Both are right, each from his or her own perspective, and a third-party observer may in turn agree with either Bob or Anna.[9] So they disagree, as indicated by Anna's reply, yet both of them may be right. There is leeway for some subtleties in the way assessors are assigned to CEs which may have applications in moral philosophy. Neither the speaker nor the hearer need to be the proper assessors, for instance, they could be ideal observers instead and many more refinements are possible.

3. *Pragmatic and Metalinguistic Replies*

One criticism of direct disagreement is that it is content-based and that disagreement is way more flexible than that. Two people may also be said to disagree if they have differing attitudes towards information that is conveyed pragmatically. For example, Lopez de Sá has argued that speakers may disagree in a contextualist semantics even if they associate different semantic contents with the same utterance, because at the same time they might pragmatically convey presuppositions of commonality.[10] In such a theory, when it is also based on speaker-contextualism, (1–a) could be understood as expressing the proposition that democracy is good from the perspective of Anna, whereas (1–c) would express the proposition that democracy is not good from the perspective of Bob, which is compatible with Anna's assertion, and they might still disagree at a pragmatic level, because each of them presupposes something like 'What is good for me, is (usually) good for everyone' and 'What is not good for me, is (usually) not good for anyone else either.' This peculiar type of contextualism has probably never been defended for goodness in general, but when 'democracy' is replaced with

---

[9] If the lack of a truth value is also allowed, then a third party observer need not even agree with any of them.

[10] See de Sá (2008, 2009, 2015).

'licorice' the natural reading of 'good' in this context may act like a surrogate of 'tasty' and for this reading the semantics may not seem less plausible than the corresponding relativist position according to which the content of Anna's utterance is 'Licorice is good' simpliciter, which turns out to be true when assessed by Anna and false when assessed by Bob. To cut a long story short, while there may always be a battling of intuitions about the question of whether and in which way various sorts of expressions and their readings are context-sensitive, de Sá's critique seems justified in general. It seems inadequate to presuppose a notion of disagreement that is only defined on the basis of what and how utterances express semantic content. This critique extends to the other content-based notions of disagreement mentioned above, because they are based on aspects of the semantic representations rather than attitudes of speakers about speech act content or other facts about the conversational situation or their own states of beliefs and desires that need not be directly connected to semantic representations.

A bigger problem with direct content-based analyses of value disagreement is that they link the expressions to certain metaethical positions like non-ideal appraiser-subjectivism that may or may not be appropriate for a given evaluative predicate. So even if they explain value disagreement correctly for certain expressions like predicates of personal taste, they may not be adequate for explaining value disagreement in general.

While authors like de Sá have tried to defend their own contextualist positions against relativism, Plunkett and Sundell (2013) have argued that many cases of disagreement are metalinguistic, yet can be substantive, are worth having, and are not merely verbal disputes. A metalinguistic approach is appealing for the analysis of evaluative expressions at first sight, because it allows one to explain substantive value disputes without committing to a contextualist or relativist semantics. Consider example (1) again. After at least 2,500 years of philosophical thinking about goodness, nothing can be said about the concept without taking sides, of course, but it is fair to say at least that many philosophers are willing to follow von Wright (1963) in his assessment that there are many varieties of goodness and that there is a reading of 'good' in (1) that is neither hedonic nor purely instrumental. As already suggested, under such a reading it is hard to justify a contextual or relativist semantics, for this would mean endorsing some naive form of speaker- or appraiser-subjectivism that many philosophers would consider problematic, and BPS deliver a more neutral analysis of corresponding types of value disagreement. Before turning to my own proposal, I therefore wish to briefly discuss their approach. Since their articles address many issues in semantics and metaethics at once, only the central points can be repeated here.

First, BPS lay out what they call 'canonical disputes'. Two persons A and B are in a canonical dispute if A's utterance expresses some se-

mantic 'object' *p* and B's utterance another semantic object *q* such that *p* and *q* are fundamentally in conflict with each other (Ibid.: 9, paraphrased). By formulating their conditions in this general way, BPS allow canonical disputes on the basis of incompatible plans and desires, so their notion of disagreement is already broader (and, consequently, less revealing) than the ones laid out above. However, they still have similar worries as de Sá about relying solely on a semantic content-based definition of disagreement and therefore formulate the following criterion:

> *Disagreement Requires Conflict in Content* (DRCC): If two subjects A and B disagree with each other, then there are some objects p and q (propositions, plans, etc.) such that A accepts p and B accepts q, and p is such that the demands placed on a subject in virtue of accepting it are rationally incompatible with the demands placed on a subject in virtue of accepting q. (BPS 2013: 11)[11]

They then argue that there are examples that intuitively count as disagreements and fulfill DRCC, even though they are not canonical in the above sense, and conclude that '...theorists take a wrong turn as soon as they conflate the question of whether a disagreement is *genuine* with the question of how the information on which a disagreement centers happens to be communicated.' (ibid: 13) This paves the way for their thesis that many substantial disagreements about utterances involving evaluative expressions are metalinguistic in nature and that they are nevertheless worth having. Here are their main examples:[12]

(4) a. That chilly is spicy!
    b. No, it's not spicy at all!

(5) a. Secretariat is an athlete.
    b. Secretariat is not an athlete.[13]

(6) a. Waterboarding is torture.
    b. Waterboarding is not torture.

(7) a. Lying with the aim of promoting human happiness is sometimes morally right. In fact it often is!
    b. No, you are wrong. It is never morally right to lie in order to promote human happiness.

(8) a. Tomato is a fruit.
    b. No, tomato is not a fruit.

According to BPS, all of the above dialogues are examples of *metalinguistic negotiation*. They concern the question of how to best use words

---

[11] This requirement fixes deficiencies of earlier definitions of disagreement based on attitudes that one person cannot hold jointly like in Egan (2010: 278). Cf. Marques (2014) for a critique of such definitions. I do not wish to enter this methodological debate here and consider DRCC a reasonable rule of thumb.

[12] See (Plunkett and Sundell, 2013: 15, 16, 19, 20, 22).

[13] Two sports reporters are discussing a horse in a race. One is calling it an athlete, whereas the others point is that only humans can be athletes. This example is originally from Ludlow (2008).

in a given context and nevertheless may be important and substantive. BPS call the business of discussing and determining how we *should* use words or concepts 'conceptual ethics'.[14] For example, in a biology class (8–a) might be the appropriate position, whereas (8–b) may be more adequate for a chef. Likewise, there is no doubt that (6) is morally relevant and a dispute worth having, even when underlying the dispute is, according to BPS, the fact that both discourse participants disagree about the meaning or adequate definition of 'torture'. What constitutes the concept under discussion is important because of '...sociological facts about its sociological role' (BPS 2013: 25), because there is something '...substantive at stake in how the relevant terms are used in the context [...] and the speakers recognize this fact.' (ibid.) As they lay out, these disputes also survive paraphrasing, a test devised by Chalmers (2011) to distinguish substantive from merely verbal disputes, so metalinguistic disagreement need not be merely verbal. That is, in a nutshell, their position.

To summarize, the problem is how to explain value disagreement in a way that does not make it a mere fighting about words and without stipulating a particular theory of value. Notwithstanding the possibility that many evaluative predicates are context-sensitive in some other ways—in fact, most of them are—, explaining examples like (1) and (6) in the contextualist way does not seem to make them cases of genuine disagreement. de Sá (2008, 2015) argues that this apparent deficiency can be fixed by explaining the disagreement by different attitudes about pragmatically conveyed content. However, relativists are not satisfied with this solution, as the contextualist approach looks like an attempt to explain away the strong intuition that the discourse participants in such examples really do talk about the same subject matter. But the relativist position presupposes a type of appraiser-subjectivism that is even stronger than contextualism, and both positions seem implausible as a general way of explaining examples like (1)–(8). They may be adequate for *some* expressions such as 'tasty' and 'fun' but seem to be unacceptable for other, broader value predicates like 'brilliant', 'right', 'is torture' and readings of 'good' that are not purely hedonist or instrumental.[15] The problem is that sometimes people might even disagree fundamentally about the underlying value theory, for example in (7) the speaker might be a consequentialist and the hearer a Kantian. Following de Sá in the assessment that the discussion between contextualists and relativists was based on a too narrow semantic view

---

[14] See (ibid: 3) and Burgess (2013); Burgess and Plunkett (2013).

[15] To avoid misunderstandings, it must be stressed that relativists generally do not deny this and also disagree among themselves about the question which expressions have a relativist semantics. In the same vein, it is also not claimed in the next section that a dual aspect semantics invariably means that no expression has a relativist semantics. Some might very well have a meaning that is best analyzed in relativist terms.

of disagreement, BPS suggest that many cases of value disagreement are potentially important and worthwhile metalinguistic negotiations.

4. *Two Faces of Meaning*

This section consists of a negative and a positive part. In the negative part, I argue that the metalinguistic approach is not satisfying as an explanation of the value disputes under consideration. In the positive part, I sketch a dual aspect semantics that fixes these shortcomings and can explain disagreement in a way that is content-based but nevertheless remains compatible with the thesis that many value disputes are about the meaning of the expressions involved. This suggestion is then defended and motivated in more detail in Section 5.

4.1. *Why Metalinguistic Negotiation Does Not Suffice*

Overall, the metalinguistic approach has many merits, as it paints a more realistic picture of some value disputes than purely semantic notions of disagreement. However, the problem is that it leaves the question open of how to define metalinguistic negotiation and it turns out at a closer look that it is very hard to make sense of this process. According to BPS, metalinguistic disagreement hinges on the idea that '...certain words (largely independent of which specific concept they express) fill specific and important functional roles in our practices[.]' (Plunkett and Sundell, 2013: 20) This passage seems to suggest that the functional role of the expression in question does not depend substantially on the concept it expresses, but is this really plausible? To me it is not, for it seems hard to find a way in which a social practice with regards to a term may come into being without being based on a widely accepted meaning of that term, or in other words, *because* the term has that specific meaning and not another one. At other places, BPS explain that two discourse participants who disagree '...each advocate a view about which concept is best suited to play a certain functional role in thought and practice...' (ibid.: 21), and so it seems that the disagreement is about the concepts again. Perhaps the question whether the disagreement is about the use of the words or the concepts is not so important, because according to BPS both come as a package. Anna and Bob both try to advocate their concept in association with a given word, arguing that the respective concept better suits or fits the functional social role of past uses of the word. The problem with this suggestion is, however, that BPS do not provide a satisfying account of the '...social, historical, and psychological facts about what is standardly associated with the use of that term.' (ibid.) To see why this is a problem, consider the following joking conversation:

(9) a. Anna: Work is torture.
     b. Bob: Indeed, it is.

They are joking, but why? The correct answer is in my opinion that they both know that in modern democratic societies without slave labor camps work does not classify as torture, therefore Anna's utterance is obviously false, and then a Gricean recovery strategy kicks in.[16] This can only work if 'torture' has enough of a partial pre-established meaning on the basis upon which Anna and Bob can agree that work does definitely not qualify as torture under normal circumstances, even if they disagree about other torture-related issues such as (6). For the dialogue to work as a joke, a pre-established meaning is responsible and *not* the fact that 'is torture' is not commonly used in a way that an utterance of 'Work is torture' is accepted. Otherwise a value disagreement like (1) or (6) would boil down to an argument not about what constitutes torture, but about how people in the past used the word—but past utterances of (6–b) might not have existed at some point in time or they might have been made by people who clearly agreed all the time that (6–b) is false. So it seems that there is a pre-established meaning on the basis of which Anna and Bob forward their own views about torture, where one or both of them might deviate from this meaning, which past uses have expressed. If so, a dispute about past uses is in the essence a dispute about this pre-established meaning. However, it seems woefully inadequate to reconstruct Anna's and Bob's dispute as being about the question whose definition of torture best fits this pre-established meaning. *That* question should not be hard to answer and there ought not be much rational disagreement, because it is a purely factual matter. Either her suggestion fits or does not fit the prior established meaning, and the disagreement would in the end be about a linguistic matter, whether other speakers in the community use the word 'torture' in a way that is compatible with, suggests, or even implies that waterboarding is torture. There is nothing to negotiate then, they could just ask other speakers, so this cannot be the right approach either.

Perhaps it helps to understand the social functional role in a very broad sense and to assume that Anna and Bob disagree about a case that is not settled by prior uses. In this view, they seek to *extend* the meaning of 'torture' in a way that harmonizes with prior social practices concerning torture such as banning it, persecuting it as a crime, and so forth. However, this cannot be right either. Even if everybody had always agreed that there is no torture in some speaker community, it is hard to see how this fact alone could settle the question or even just play a substantial role in determining whether waterboarding is torture. Suppose Anna and Bob continue to discuss (6) and finally come to agree that waterboarding is torture. According to theory that BPS seem to suggest, Anna would essentially have convinced Bob of the fact that claiming that waterboarding is torture best suits our previous

---

[16] This is only an example. There are other explanations such as those given by Relevance Theory and nothing in what follows hinges on assuming the classical Gricean picture.

practices of using 'torture' and reasonably fits with the already existing sociological functional role of the word. But if Anna and Bob negotiate in that way, they still do not settle the question whether waterboarding *is* torture! They could both agree that waterboarding is not torture and be wrong about it. How so, if they are just negotiating? It seems that the answer is this: They really argue about what constitutes torture in (6), about what makes socio-economic systems good or bad in (1), and about what's morally right in (7), and in each case they argue on the basis of some already shared, pre-established meaning but without necessarily presuming that this meaning is correct or adequate.

This critique does not concern the metalinguistic aspect of the proposal. Dialogues like (1)–(8) are metalinguistic in the sense that they concern the meaning of expressions involved, and it seems clear enough that these expressions are not used in the ordinary way in these dialogues. I fully agree with BPS about this observation. The critique challenges the view, however, that Anna and Bob negotiate the best or most appropriate use of a term. The thesis that in (6) Anna suggests to Bob that waterboarding should be called 'torture' because of existing practices associated with the word is not very convincing as a *general* explanation of such disputes, because this would make her use of 'torture' a mere proxy or surrogate for something else. Instead, the correct inference goes the opposite direction: Since the word expresses a certain concept, it is endowed with a certain social function, and another use of the word is endowed with the same (or a more specific) social function *insofar as* whatever it expresses falls under this concept. Hopefully, in Anna's society torture is morally prohibited, illegal, and socially unacceptable, and she may also use (6–a) to implicate or otherwise pragmatically convey these facts with the intention of laying out to Bob that waterboarding should be illegal, socially unacceptable and prohibited. But even if she does so, a use of (6–a) to successfully convey this additional speech act content can only be based on the prior meanings of the words in the utterance.[17] If waterboarding is not torture in the first place, as Bob believes, then how could (6) serve to indicate or even justify that waterboarding should be illegal? Alice's reasoning chain must thus go in the following direction: Waterboarding is illegal, socially unacceptable, and so forth, because it really *is* torture, and what is considered torture in her society is considered illegal, unacceptable, and so forth. If that is the right direction of explanation, then the conundrum remains: If, on the one hand, that meaning is the pre-established meaning, then the dispute is factual, as laid out above. If, on the other hand, it is Anna's meaning as opposed to Bob's favorite definition, then the disagreement is once more direct and content-based, hence the two of them are talking past each other and the relativist critique applies.

---

[17] Radical pragmatists and meaning eliminitavists would not subscribe to this view. These positions discard with truth-conditional semantics altogether and cannot be taken into consideration here for lack of space.

Similar points can be made about the other examples. However, these considerations remain compatible with the view that metalinguistic negotiation sometimes takes place in dialogues like (1)–(8), they merely indicate that BPS's approach is not adequate as a *general* explanation of these kinds of examples. Doubts remain about their use of the term 'negotiation', which seems to be too weak for an explanation of at least some cases of important value disagreement, namely those in which the social function of an expression arises or is even constituted from its meaning. The role of negotiation in the metalinguistic approach needs to be motivated further. As it stands, there is a gap between the pre-established meaning of an expression and its social function. If the latter is determined or significantly constituted by the former, then one of the disagreeing discourse participants may either be right by virtue of this meaning, if the new use is 'covered' by it, and the other will be wrong (or, both are wrong). This is a factual matter. The second possibility is that the new use is not covered by the prior meaning, as could be argued for the 'athlete' example (5), for instance, but then the extension of the new meaning is not covered by the social function either and the existing social function of the expression alone does not generally provide a reason for extending the prior meaning to the new use. BPS are right in calling the process of arguing for such an extension negotiation, as this process is oddly idle: The evaluative predicate has no power of its own, its existing truth-conditions barely matter, and it merely seems to act as a surrogate for attributing properties of the social function to the logical subject of an assertion. This is implausible. Although there may be exceptions to this rule, generally speaking linguistic expressions have a social role because of their meaning, not their meaning because of an existing social role.

None of the above implies that metalinguistic disagreement cannot sometimes occur, and BPS have argued convincingly that at least some such disagreements can be substantial and worth having. The following dual aspect approach should therefore be taken as an addition to their account, as a way of making it more precise.

4.2. *Core Meaning and Truth Conditions*

To address the above mentioned shortcomings semantic theory itself needs be modified. This makes the proposal more controversial, of course, but it also becomes more expressive and the modified theory also explains many more phenomena than value disagreement. In the proposed dual aspect theory (DAT), every expression of a language has two meanings. The *noumenal meaning* of an expression contributes to the truth-conditions of a sentence as a whole in accordance with Wittgenstein's dictum. In contrast to this, the *core meaning* of an expression represents what I have so far called pre-established meaning.[18]

[18] There is an older use of 'dual aspect theory' in the philosophy of language, which is based on the idea of enriching meaning representations of sentences

This meaning is best understood by a loose bundle view, it represents the smallest common denominator between speakers of a given sociolect insofar as the meaning of the given expression is concerned, and speakers of a linguistic community have to agree to a large extent on this meaning for communication to succeed and in order to count as a competent speaker. Both meanings may be truth-conditional, because the attribute 'truth-conditional' marks a *method of specifying certain aspects of meaning*. Moreover, both types of meaning may be spelled out in contextualist terms in the way laid out above to deal with indexicals and other forms of overt context-sensitivity, and this will be assumed from now on without further mention.

There is a third element that for lack of a better term shall be called *noumenon*. It is the (stipulated) object of an intentional stance of speakers towards reality; noumenal meanings aim at the noumenon, they may pick out objects in reality like Fregean senses aim at the 'Bedeutung', and sometimes reality also *informs* speakers to revise the noumenal meaning of an expression. Noumenal meaning is also a précis of the core meaning, whereas the core meaning is first and foremost what is needed in everyday conversations. Core meaning evolves primarily out of a need to communicate and coordinate behavior, it may be conventional in the sense of Lewis (1969) or may be more broadly conceived conventional in the sense of being based on non-inferential behavioral regularities between word and meaning. Noumenal meaning generally evolves out of core meaning when people start asking questions that aim at reality. For example, two speakers can talk about a lightning during a thunderstorm and agree that it needs to be avoided and that it is likely that fire can be found where the lightning has struck. But they may also start asking themselves what a lightning really is. Is it a sign from the gods, some special form of fire, an electromagnetic phenomenon? These questions concern the noumenal meaning.

Core meaning is often incomplete and loose. For example, 'freedom' is an abstract noun that in everyday use stands for the possibility of making voluntary choices, an absence of oppression and unnecessary prohibitions, a lack of restrictions in thinking and acting, et cetera, but different speakers will only loosely agree on such features. In contrast to this, the scientific definition of an expression by an expert in the respective field is a possible candidate for a noumenal meaning or an approximation thereof. Kant's definition of 'freedom' is such a candi-

---

occurring in attitude ascriptions with additional features (guises, cognitive roles, ways of being given, Fregean senses, etc.) in order to explain cases of referential opacity. For example, Lois Lane might believe that Superman can fly while at the same time consistently believing that Clark Kent cannot fly, because Superman is given to her in cognition in a different way than Clark Kent. Core meaning does not fulfill this role, it serves the opposite goal of standing for the lowest common denominator and is therefore looser than noumenal meaning rather than more fine-grained. The two theories have nothing particular in common. To combine them, a triple aspect theory would be needed.

date, for example. Importantly, both meanings may be revised on the basis of a change to the other. The third element, the noumenon, is not a sort of meaning but something else in reality such as a mathematical fact, the orbit of the Earth around the sun, a particular family of cats, a building, or a person, if the noumenon exists at all. Since it is not a sort of meaning, the account is a dual rather than a triple aspect theory.

Before defending DAT in more detail, let us take a look at how it fares with the problematic examples. When two speakers disagree like in (1)–(8), they (mostly) agree about the core meaning and at the same time disagree about the noumenal meaning. They also *intend* the noumenal meaning to match some aspect of reality.[19] So what they really disagree about is the question whether *their* respective candidate for a noumenal meaning is adequate with respect to the stipulated noumenon and can therefore be regarded as a précis of the imprecise everyday core meaning on which they both implicitly agree. If one of the discourse participants does not associate a core meaning with an expression that is sufficiently similar to that of other members of the linguistic community, as they evidence from that speaker's odd linguistic behavior, then they will raise doubts about his competence as a speaker of the respective sociolect. Core meaning is linguistically mandated and the basis of a shared lexicon. For example, it is part of the core meaning of 'torture' that it involves inflicting serious harm (often though not always in an attempt to extract information), it is part of the core meaning of 'athlete' that an athlete is physically fit under normal circumstances, and so on. If Bob claimed that torture does not harm the person who is tortured, then he would not be considered a competent speaker with respect to 'torture' and Anna would react accordingly, perhaps by shaking her head in disbelief or by trying to educate him about the English language. In contrast to this, competent speakers may disagree about the noumenal meaning at any time, it is principally contestable and often hinges on complicated background theories which might turn out to be false. Disagreement about noumenal meaning is never about this meaning alone, though, which would lead to similar problems as those laid out for metalinguistic negotiation in the previous section, but rather about the question whether this meaning adequately captures the respective aspect of reality it is intended to capture—whether it correctly points to the noumenon.

From this perspective, the disagreement between Anna and Bob is metalinguistic insofar as it concerns an aspect of the meaning of expressions used, but it is also linguistic and a common form of disagreement, because it is based on a mechanism that is completely normal for language users and occurs daily. Linguistic competence does not require convergence on noumenal meaning, on the contrary we often discuss it when we talk about general terms that are not grounded in

---

[19] Whenever intentions are mentioned in this article, they are to be understood in the philosophical sense that started with Brentano, not as intentions to act.

empirical facts in any obvious way. In such discussions, expressions are not mentioned in the sense of being quoted, but the underlying concepts are challenged by one or more discourse participants when they advocate some particular noumenal truth-conditions for the term. This may, but need not be an argument for or against a relativist or contextualist semantics, as (3) and (4) might exemplify, because every aspect of the noumenal meaning is open for discussion. Some speaker might argue for the assessor-sensitivity of a given evaluative predicate, for example. A discussion may also amount to a mostly verbal dispute, as one might consider Ludlow's 'athlete' example (5), for it is questionable whether there is something in reality that can decide whether horses can be athletes or not other than the existing core meaning or arbitrary conventions.[20] In yet other cases such as (8) the dispute might be best characterized as a linguistic misunderstanding of confusing contextual standards, and this is also how BPS discuss this example. In a biology class the noumenal meaning is decisive, since biologists study plants and know criteria for classifying tomatoes that ordinary speakers might not even have heard of. If Bob is not aware of this context and defends the core meaning, he is right, too, but in a sense that is often accompanied with a sigh. He is wrong in the same sense as a student in a philosophy class would be ill-advised to use 'realist' in the sense of 'an uncomplicated and reasonable person, being down to earth'.

Some such cases might be conflicts between different sociolects, each equipped with its own noumenal and core meaning for a term, while others may concern the context of use, whether the word is used in an every day sense, or whether we are interested in *what it really means*. But this inaccuracy is not a deficiency of DAT, it merely reflects the reality of disagreements that everybody has experienced before. What is important for my thesis is that in clear cases of indirect value disagreement like (1), (6) and (7) the dispute will normally be about the noumenal meaning *on the basis of the idea that something in reality corresponds to this meaning and thereby justifies it*. Provided that both suggestions are sufficiently compatible with the core meaning established by the linguistic community, such a dispute will be merely verbal and insubstantial only if (a) competent speakers do not intend the noumenal meaning to 'fit' the noumenon *and* (b) there is nothing in reality that would justify the meaning. Notice that error theorists defend (b) but need not deny (a). As Meinong (1971) argued pervasively, having an intention in the sense of the Brentano School need not imply that the intentional object exists.[21]

---

[20] Bear in mind that many disputes may also be about the core meaning; speakers are not always a hundred percent competent, the boundaries between sociolects can be vague, and it may be disputed which sociolect is relevant to the discussion.

[21] I have laid out my own stance about nonexistent objects in Rast (2011). However, abstract objects are more complicated, see Zalta (1983) for a logically and philosophically developed approach.

Regarding value disagreement, the theory explains how speakers may disagree about the meaning of expressions without talking past each other in the sense of direct contextualist disagreement and without assuming a relativist stance that may only be adequate for select predicates of personal taste. Principle DRCC is fulfilled, because the speakers disagree about the noumenal meaning of an expression on the basis of a given core meaning *and* and on the basis of their beliefs about the noumenon. The theory is therefore an extension of the metalinguistic approach, which on its own has problems with explaining why and what people are negotiating about the meaning of an expression. However, if this was the only merit of DAT, then the metalinguistic approach would perhaps be more desirable, since any unnecessary duplication of entities should be avoided by application of Occam's Razor. As I will argue in the next section, DAT has many additional benefits and is nearly unavoidable for semantic externalists anyway.

5. *In Defense of DAT*

The dual aspect approach laid out in the previous section is not new. Similar suggestions have been made by semantic externalists and 'anti-individualists' like Putnam (1975b) and Burge (1979, 1986) on the basis of insights from Kripke (1972). However, some care is needed to keep different aspects of externalism apart, since otherwise we would only win a pyrrhic victory. DAT aims at remaining compatible with weak semantic internalism by focusing on the social aspect of linguistic labor division and allowing at the same time that two disagreeing parties may have different noumenal meanings *in mind*. In contrast to this, the strong externalism of Putnam (1975b) stipulates that a description of the extension constitutes part of the meaning of a term and that knowledge of this description or of the extension itself are not required by virtue of linguistic competence. The problem with this view is that it seems fairly implausible as a requirement for moral terms, because it presupposes an implausible form of moral realism, and this view would also be incompatible with more recent work in Putnam (2002). I will argue for a weaker form of externalism that explains metalinguistic disputes by the fact that speakers disagree about the noumenal meaning of expressions and in order to satisfy DRCC speakers must somehow be aware of these meanings in their minds in case of a genuine disagreement. This weak externalism for value terms remains compatible with the strong externalism of natural kind terms and nothing in what follows implies that strong externalism needs to be given up for those terms.

5.1. *Arguments from Social Externalism*

All of the arguments for semantic externalism cannot be repeated here for lack of space. However, I want to sketch two of them very briefly in

order to carve out the distinction between weak and strong externalism
I am aiming at. There are two types of arguments for semantic exter-
nalism. The first one is based on Twin Earth scenarios, whereas the
second one pertains to social labor division and is empirical. According
to the empirical thesis, individual speakers of a natural language like
English may generally be judged as competent speakers, even though
they do not, according to the judgment of experts, 'associate' or 'grasp'
the truth-conditional contribution of an expression to a sentence-in-use
as a whole precisely enough.[22] For example, speakers of English can use
'birch' competently without being able to distinguish birches from elms,
can use 'water' without knowing that it is H_2O, and can use 'capitalism'
competently without implicitly knowing or being able to successfully
apply a definition that would satisfy an economist. The empirical evi-
dence for social externalism is pervasive and there can hardly be any
doubt that there is a linguistic labor division. We leave the question of
what exactly certain expressions mean to experts and generally assume
that these experts have a precise idea about what these expressions
mean. Empirical arguments for social externalism are compatible with
semantic internalism, though they certainly make it less appealing. An
internalist could claim that at least some expert has to implicitly know
the 'right' meaning of an expression in order for it to make sense at all;
if that were not so, the internalist could argue, sentences containing
such expressions might not be about any common subject matter at all.

Putnam's Twin Earth argument closes this loophole. Suppose John
on Earth has Twin John on Twin Earth as a counterpart who is mol-
ecule-to-molecule identical to John and also identical in his mental
life with him, except that he lives on Twin Earth where XYZ replaces
H_2O.[23] Putnam argues that even though they are in identical mental
and physical states, in this scenario 'water' uttered by John on Earth
refers to H_2O, whereas 'water' uttered by Twin John on Twin Earth re-
fers to XYZ. Hence, the meaning of 'water' is not in their heads. Similar
arguments can be made for other natural kinds like 'tiger' and 'gold'.
These arguments are clearly not empirical, since otherwise the exis-
tence of Twin Earth and XYZ would have to be established for the ar-
gument to be sound. Instead, they are thought to reveal the principally
indexical nature of certain expressions: The extension of 'water' is fixed
indexically and possibly independent of our current state of knowledge.
If the word is used on Earth, then it refers to H_2O. If it is used on Twin
Earth, then it refers to XYZ.

---

[22] The scare quotes around 'associate' and 'grasp' are certainly adequate in
this context. Ideally, much more would be said about these problematic metaphors
but for lack of space this issue has to be suppressed. I will continue using the verb
'associate', which may be taken to stand for a form of implicit knowledge or some
ability.

[23] It is sometimes argued that John and Twin John could not be identical, because
the latter partly consists of XYZ instead of H_2O. While technically correct, this
argument leads astray and ultimately does not invalidate the thought experiment.

Since it would be implausible to stipulate that speakers of a language must know the correct extension of each term and this would also conflict the empirical observation of social externalism, and since the idea that a systematic error-theory according to which speakers are generally not competent is equally unappealing, Putnam suggests to represent the meaning of general terms like 'water' or 'tiger' by a vector that contains semantic and syntactic markers such as 'count noun' and 'natural kind term', the specification of a stereotype that comes close to what I have called core meaning, and a description of the term's extension, i.e., H_2O for 'water' on Earth, but XYZ for 'water' on Twin Earth.[24] A competent speaker must only sufficiently master the first two items in a meaning vector, whereas the description of the extension is relevant for the externalist determination of meaning that Twin Earth scenarios are supposed to establish.

Putnam's position is very close to the proposed dual aspect theory. The core meaning may be taken as being comprised of the stereotype and semantic and syntactic markers, whereas what Putnam calls a 'description of the extension' corresponds to noumenal meaning. However, it differs from the current proposal with respect to the role of an extension (if there is one). Putnam (1994) meanders between different readings of the third elements in his meaning vectors; he either takes it to be a *description* of the extension or as the extension itself.[25] In contrast to this, I would like noumenal meaning to be more broadly understood as a refined meaning with which informed speakers come up when they are intending to narrow down and state more precisely a purported aspect of reality that is *suggested* by a preestablished core meaning and sometimes, though not always, also *perceived* or *measured* from observing nature. An important difference between these views is that in DAT the core meaning of predicates must enable the language user to identify entities falling under them—or, in case of non-empirical general terms such as 'triangle' and 'democracy', to identify instances of the reified abstract objects they express—in a sufficiently precise way whenever the noumenal meaning is under discussion. This process of identification need not be foolproof, but it needs to be precise enough to ensure that two people roughly talk about the same issue. Take 'democracy', for instance. While the concept of democracy is characterized to some extent by examples, I submit that it cannot be defined in a purely ostensive manner. But in order for two discourse participants to discuss democracy, according to DAT they need to roughly agree on relevant aspects of the core meaning prior to being able to successfully discuss the noumenal meaning, and this involves being able to tell a democratic from an obviously totalitarian society. This prior identifica-

[24] See Putnam (1994: 252–255; 270–2), cf. Putnam (1975a: 148–152). Accounts of the nearness of semantic representations to each other and of individuals falling under them from a prototype have been made precise in geometrical meaning theories by Rosch (1975, 1983), Gärdenfors (2000) and others.

[25] See Putnam (1994: 270–1).

tion step is missing in Putnam's stereotype theory and even explicitly denied as a requirement for empirical terms, but it is crucial for understanding value disputes. According to DAT two speakers are talking past each other and cannot disagree about the noumenal meaning in any fruitful way as long as they do not have at least a glimpse about the noumenon already.

5.2. *The Peculiarities of Value Disagreement*

Externalism only *indirectly* supports the explanation of value disagreement that was tentatively outlined in the previous section. As BPS point out convincingly, it would be implausible to claim that experts generally fix the meaning of value terms, as there is persistent disagreement among experts about the meaning of particular uses of 'good' and other evaluative predicates.[26] Some forms of goodness are almost purely factual, however, and for these social externalism seems plausible. To this category belong medical goodness (healthier than, better with respect to health) and varieties of instrumental goodness such as what constitutes a good hunting knife or a good telescope. For a given purpose, such questions can be answered by experts, and these notions of goodness are based on criteria upon which experts concur at least to some extent. It is doubtful, however, whether the question of what constitutes a good life can be answered authoritatively by an expert, there is no agreement between those experts, and there is even disagreement about who counts as an expert in this domain. Similar doubts are in place about moral goodness, since otherwise skeptical arguments like that of Mackie (1977) could not have gained traction. So social externalism is only plausible for some limited forms of instrumental goodness, and the claim that these constitute other forms of goodness can only be made from an already assumed perspective, for example from the perspective of consequentialist utilitarianism, and not in general.

An even bigger problem occurs when strong externalism of value terms is presumed. This position is inherently problematic for evaluative language because it seems to presuppose value realism. Perhaps we point out instances of good things and actions, somehow access real values by these practices, and thereby define what a particular reading of 'better than' means, but regardless of how plausible one finds this theory, it remains just one of many possible explanations of the meaning of value terms. Lexical semantics cannot decide which moral theory is correct. Strong externalism works for empirical terms, because the assumption of metaphysical realism, the assumption that we all perceive the same actual world that is such-and-such, is philosophically acceptable to many if not most contemporary thinkers. Assuming a similarly strong realism for value terms is way more problematic.

[26] See Plunkett and Sundell (2013: 26–8).

Firstly, it would exclude a variety of metaethical positions, and secondly, strong arguments have been devised against forms of value realism that are based on strong externalism.[27]

So there are tensions between arguments for semantic externalism and plausible views about the meaning of value terms. The support is indirect, since both types of arguments for externalism count in favor of DAT, which in turn explains the problematic cases of value disagreement discussed in the previous section. In order to make this story convincing, though, more has to be said about the way in which DAT explains value disagreement. To see how DAT fares in comparison, consider the moral dispute in (6) again.

Why do the discourse participants continue to disagree at all, why can't they just stipulate that waterboarding is torture$_1$ but is not torture$_2$ and happily agree with each other thereafter? According to the story of BPS, this is so, because of existing practices associated with 'torture' that are morally relevant such as condemning it, making it illegal, and so forth. However, this story is incomplete, because the discourse participants could continue to discuss whether these practices should be associated with 'torture$_1$' and 'torture$_2$' respectively, and nothing of value, it seems, would have been lost by making this terminological clarification. Why would they need to negotiate one single use and meaning of 'torture'? DAT answers this question and thus provides the missing link. The discourse participants are arguing about the noumenal meaning of 'torture', what *really* constitutes torture, on the basis of a prior partial agreement about a core meaning, which involves criteria like causing substantial harm to the victims. They disagree about the right way of making these potentially incomplete and loose criteria precise. Should psychological harm be included? How much of it? What constitutes substantial harm? What do both moral and legal experts say about it? What do international humanitarian conventions and human rights say? These questions have been discussed for some time.[28] From the perspective of DAT, the speakers do not introduce mere definitions for different sets of criteria, because the noumenal meaning of

---

[27] Horgan and Timmons (1992) devise a Moral Twin Earth scenario. If on Earth 'good ' has a consequentialist meaning and on Twin Earth it has a deontological meaning, then we would not say that 'good' on Twin Earth and 'good' on Earth have different meanings like in the case of water, but according to Horgan & Timmons we would say that earthlings and twin earthlings have different theories of goodness (and likewise for 'right'). One need not share their intuitions and may also have doubts about the way they formulate semantic positions like Putnam's, but in any case a strong externalist of moral terms has to address this argument. In contrast to this, from the perspective of DAT earthlings and twin earthlings simply disagree about noumenal meanings and there is nothing special about the scenario that would set it apart from other disagreements about value terms.

[28] It may be argued convincingly, though, that the 'pro-waterboarding' side has always considered the practice torture and just flat-out lies for the purpose of strategic maneuvering. So perhaps this is not the best example, but I'd like to stick with it because BPS use it.

'torture' is intended to capture some aspect of reality just like the word 'electron' does—social, human, legal and moral reality in the former case, physical reality in the latter.

The use of 'good' in (1) is slightly different than the waterboarding example and indicates a feature that might be peculiar to value disagreement. Suppose Anna has an understanding of 'good' in mind that could be paraphrased as 'good for me and my business', since she is a business owner and does not aim at comparing different conceptions of society. Suppose that Bob understands it differently as 'good for society as a whole' and happens to be a die-hard communist. The criteria they associate with the words could not be more different, yet Anna might actually *agree* with Bob about his assessment. She just has another reading of 'good' in mind. The difference to the previous example is in this case that 'good' and its more fundamental comparative counterpart 'better than' are *also* context-dependent and have many different noumenal meanings as acceptable précis of the same underlying core meaning in different contexts. So in addition to disagreement about a particular noumenal meaning, competent speakers may also disagree about the way in which a use of these predicates ought to be understood within a given conversational context even if they agree on the available readings. This might be a peculiarity of what I call *aggregative value expressions* to which 'good' belongs. These are multi-dimensional expressions that combine various evaluative aspects of two or more items under discussion into an overall evaluation. Although I do not wish to make the case for this position here, I do believe that all genuine value expressions are aggregative in this sense.[29] In contrast to this, we could call 'torture' value-laden, because the vast majority of mentally sane persons consider it to be bad, but at the same time might not consider it a value term itself. Perhaps this is the reason why we say that torture is bad but do not often explicitly state that being healthy is good or that being brave is good: The adjectives 'healthy' and 'brave' are themselves value terms that take part of an analysis of respective medical and instrumental varieties of goodness. If this is the right view, then we may expect candidates for the noumenal meaning of value terms to be fairly complex and subject to particularly long-lasting controversies. Pace Mackie (1977), widespread disagreement about value terms need not be explained as an inference to the best explanation that there is no noumenon. Maybe a better explanation is that the disagreement is so persistent because value terms concern complex social matters whose evaluation requires considering a vast number of factors with potential trade-offs between them.

5.3. *Core Meaning and Prescriptivism*

In his famous book on evaluative language, Hare (1952) understood his universal prescriptivism as a corrected and expanded version of the much criticized Paradox of Analysis by Moore (1903):

---

[29] Stojanovic (2015) sketches such a semantic approach. Pragmatic theories are briefly addressed further below.

> Moore thought that he could prove that there were no such defining characteristics for the word 'good' as used in morals. His argument has been assailed since he compounded it; and it is certainly true that the formulation of it was at fault. But it seems to me that Moore's argument was not merely plausible; it rests, albeit insecurely, upon a secure foundation; [...] Let us, therefore, try to restate Moore's argument in a way which makes it clear why 'naturalism' is untenable[.] (Hare 1999: 83–4; Sec. II.5.4)

Among Hare's arguments, the cannibals/missionary thought experiment deserves special mention, not only because it is discussed by BPS and others, but for our purpose also in order to show why Hare's preskriptivism and similar positions are fully compatible with DAT. He writes:

> Let us suppose that a missionary, armed with a grammar book, lands on cannibal island. The vocabulary of his grammar book gives him the equivalent, in the cannibals' language, of the English word 'good'. Let us suppose that, by queer coincidence, the word is 'good'. And let us suppose, also, that it really is the equivalent—that is, as the *Oxford English Dictionary* puts it, 'the most general adjective of commendation' in their language. If the missionary has mastered his vocabulary, he can, *so long as he uses the word evaluatively and not descriptivively*, communicate with them about morals quite happily. They know that when he uses the word he is commending the person or object that he applies it to. The only thing they find odd is that he applies it to such unexpected people, people who are meek and gentle and do not collect large quantities of scalps: whereas they themselves are accustomed to commend people who are bold and burly and collect more scalps than the average. (Hare 1999: 148; Sec. II.9.4; orig. emph.)

This much-cited passage occurs late in the book after Hare has already made the case for the commendatory nature of value terms and suggested to analyze them as sort of hidden imperatives. Its purpose is to illustrate why the descriptive meaning of value terms, whose existence Hare does not deny, are secondary to their evaluative use as tools of commendation. The missionary and the cannibals can understand each other, because they associate the same evaluative meaning with 'good'—that of general commendation—, even though they associate opposing descriptive criteria for what constitutes a 'good person' with the word.

From the perspective of DAT, such examples can be explained by stipulating that part of the core meaning of positive value terms like 'good' is commendation, and part of the core meaning of negative value terms like 'bad' is the opposite of commendation, however one may call it. In everyday language we use positive value terms to commend things or persons with a certain respect. For example, healthy food is food that is commendable with respect to improving or maintaining one's health and a good action is one that is commendable 'all things considered'. From this point of view, competent speakers agree on this general component of the core meaning of value terms. If an honest and sincere speaker calls action A better than action B, all things considered, and still chooses action B or appraises B more than A, then

he lacks linguistic competence or makes a performative mistake. Even experts disagree, however, about what *really* makes an action commendable—and the proposed criteria in my view constitute part of the noumenal meaning of the respective use of 'better than'.

As appealing as this position might seem, I am reluctant to identify the core meaning of 'good' and related positive value terms entirely with commendation within a given domain. Hare himself restricts this analysis to *moral* uses of 'good', which he considers entirely different from other uses,[30] and there may be everyday uses of evaluative adjectives whose core meaning goes beyond mere commendation. Many aspects of goodness hinge on factual matters and at least in relatively homogeneous speaker communities people may agree on many readings of value terms way more than one might suspect at first glance. For example, phrases like 'a good hunting knife' (instrumental goodness), 'a good holiday trip' (hedonic goodness), and even 'a good life' (ethical goodness) may have core meanings that go beyond mere commendation and noumenal meanings upon which even non-experts converge upon sincere reflection. Despite some disagreement on the details, there seems to be vast agreement on what constitutes a good hunting knife or makes a good holiday trip, and maybe also negative soft constraints take part of the core meaning: A knife with a lousy handle cannot be good, a trip on which both the hotel and the weather is horrible cannot be good, and a life of stress, unhappiness and misery can hardly be called a good life. Since there is also reason for doubt that moral and non-moral uses can be distinguished clearly, even seemingly moral uses might be based on a fixed set of fuzzy criteria within a speaker community. Is hedonic goodness a moral concept? For the Kantian, it is perhaps not, but for many classical utilitarians it is, and as long as we are not talking about what is *really* good in the hedonic sense, we may agree that being good in this sense means conforming to our subjective preferences and tastes, for instance, which vary from person to person. That being said, commendation always seems to be part of the core meaning of positive value terms, and their noumenal meaning may be understood as different attempts to find criteria as to why the actions, objects, or persons that satisfy the core meaning are commendable—and from such an attempt it might very well follow that, judging from the proposed noumenal meaning, they are not commendable at all. So DAT can accommodate Hare's prescriptivist intuitions while at the same time allowing for the possibility that other philosophical positions about evaluative language may *also* be correct, by making the commendatory aspect of value terms part of their core meaning and explaining some cases of indirect value disagreement among competent speakers as disagreement about what the respective expressions really mean, viz., about their noumenal meaning.

---

[30] See Hare (1999: 140).

5.4. *Further Objections and Replies*

Despite its merits there are some objections to DAT that I will address and attempt to defuse in the following paragraphs.

A rather obvious objection is based on Hare's arguments and other variants of the Paradox of Analysis. In light of these arguments, does it even make sense to claim that value terms have a noumenal meaning? Would it not suffice to identify the evaluative meaning of value predicates as some sort of hidden imperative, as Hare suggests? In my opinion, this critique misses the point. Hare's position and that of many other noncognitivists may be described as the view that value terms are not associated with any fixed set of criteria into which they could be lexically decomposed, hence there might also be no defining characteristics that would justify talking about a noumenal meaning. Instead, these theories explain, in various competing ways, their core meaning by the ways these terms are connected with emotive responses, prescriptive uses, or their commendatory nature. But these are standpoints in moral philosophy, and from the perspective of DAT these are fine and compatible with the more modest aim of DAT to explain certain cases of value disagreement that do not seem to be based on direct, content-based disagreement. Expressed in terms of DAT, noncognitivists argue that the noumenal meaning of some or all value terms is (a) not given by descriptive features of the noumenon alone, because there are none or we do not have sufficient epistemic access to them, but by some features of the psychological or social role of those terms, or (b) that the noumenon of value terms consists of certain psychological or social facts that reflect and explain the commendable nature of their core meaning. This standpoint is compatible with DAT, and a major selling point for DAT as a semantic theory is precisely the fact that it can explain value disagreements like (7) between two speakers who look at the same issue (the noumenon) from vastly different metaethical stances.

Another, perhaps more serious objection concerns the semantics/ pragmatics distinction and has also been raised from time to time against contextualist and relativist positions in the Philosophy of Language.[31] The idea is to keep semantic representations minimal— though not 'crazy minimal' (Recanati 2006) like in Cappelen and Lepore (2004)—and resort to pragmatic explanations for the rest. For example, Väyrynen (2013) defends a pragmatic approach to thick evaluative concepts like 'lewd'. Would it not be better to bite the bullet and generally declare core meaning the 'real', primary meaning of expressions and abandon noumenal meaning altogether? This amounts to explaining disagreements like (1)–(8) in purely pragmatic terms, similar to what de Sá (2008) has suggested for the relativists' puzzles. It would go beyond the scope of this article to address this issue, especially since authors tend to presume their own semantics/pragmatics distinctions

---

[31] See for example the critique by Bach (2005) on Recanati (2004).

and different answers can be given for different types of evaluative expressions. One powerful objection to a pragmatic approach, however, is worth mentioning in the current context: If noumenal meaning is generally explained in pragmatic terms, then this seems to amount to giving up truth-conditional semantics altogether, because the core meaning of many expressions is often semantically underdetermined and does not contribute enough to the truth-conditions of an utterance. Such a theory has undesirable consequences. Many expressions do make a precise truth-conditional contribution to the whole utterance in the form of noumenal meaning on which experts generally agree, even when experts do not completely agree or when their definitions are literally false or incomplete. For many purposes, the deficiencies of natural language use can be ignored, since any good theory is by its very nature highly idealized. To give an example, taking half-integer spin as a defining characteristic of fermions and integer spin as a defining characteristic of bozons might turn out to be inadequate in the future, but is this an argument against the adequacy of these definitions *now*? I believe not. Such revisions often occur in science, for example the Ur-meter was replaced by a definition based on the speed of light and the definition of a second was changed from one based on solar days to one based on the radiation of cesium 133 atoms. Despite such differences in the detail, we can say that both definitions sufficiently approximate the noumenon and are therefore adequate to a certain degree in many contexts of use. For example, a speaker who utters 'This doorway is 2.20 meter high' may be said to fully understand the truth-conditional contribution of 'meter' in her utterance *as long as* she does not use it in a way that does not sufficiently approximate a meter. To make this clear, Wittgenstein's dictum of the beginning of this article is false, since mastery of the noumenal meaning is not required on behalf of speaker competence and core meaning may be too imprecise or unspecific to yield definite truth-conditions. But we may still say that a speaker understands an utterance in a given context if the truth-conditional core meaning sufficiently approximates the truth-conditional noumenal meaning of the expressions involved in accordance with the respective standards of precision that are in place in the given context. One might even go further and replace understanding with understanding to a certain degree, as long as it is kept in mind that for many purposes of coordinating behavior mastery of the noumenal meaning is not required. So there are two kinds of understanding in DAT. On the one hand, in successful communication a hearer may be said to fully understand an utterance without implicitly knowing its noumenal meaning. On the other hand, a speaker can be said to understand what a word or sentence in use really means when he or she grasps its noumenal meaning.[32] Whenever experts mostly agree on it,

---

[32] Neither of these types of understanding is based on speaker meaning or speech act content.

the noumenal meaning of an expression serves as a corrigendum of the core meaning; if noumenal meaning is discarded in this model, then the core meaning deteriorates into some vague notion of cognitive meaning suitable for many branches of linguistics but disconnected from truth-conditions. This is a high price to pay. Why give up truth-conditional semantics entirely if a truth-conditional DAT can explain, at the same time, the directedness of everyday language towards truth and reality, its use as a convenient tool for co-operation of behavior for which truth may sometimes only play a minor role, and the vagueness of truth-conditions of utterances in contexts of daily language use?

While I want to leave the exact nature of core meaning open for the time being, I have argued that it can be spelled out in terms of possibly semantically underdetermined truth-conditional meaning in a loose bundle view or the like, and noumenal meaning is truth-conditional just like Kaplan's character. This means that both of them are based on lexical decomposition of word meaning, which unfortunately also comes at a price. Any such account of word meaning must somehow defend itself against arguments that are directed against analyticity. Many of those arguments such as the Paradox of Analysis itself are not particularly convincing, and sufficient doubt has been cast on them elsewhere, so this section shall end with only some admittedly cursory remarks about these sort of criticisms which can be raised against many more semantic theories than just DAT. The idea behind them is generally that the lexical decomposition of a value term G into multiple criteria $C_1,...,C_n$ or complex logical combinations of criteria is implausible as an account of lexical meaning, because it would make certain value judgments analytical and this fact would fly into the face of our intuitions. Note that without the last part there would be no problem and critiques rarely attempt to justify why our intuitions are worth a penny, but let us buy into this reasoning for the sake of the argument. One might reply to it that since Quine (1964) analyticity is often conceived as coming only to a certain degree rather than giving rise to analytic judgments with apodeictic certainty, and that DAT is compatible with this point of view because it states that (a) different speakers need only to associate similar core meanings with the same expression, (b) what the expression really means is often left to experts as part of a linguistic labor division, and (c) defining characteristics of both types of meanings may be false or otherwise not suitable, as long as there remains a way to identify the noumenon on the basis of those characteristics. For instance, although very unlikely, current chemistry might turn out to be incorrect and in need of revision because water turns out to be XYZ. Suppose it is XYZ. Then the noumenal meaning upon which experts unanimously agreed is inadequate, false, incoherent, or otherwise in need of revision due to changes in the underlying background theory. So analyticity is given the role that Quine reserved for it, not as a justification of infallible judgments, but as an explanation and indicator of

our willingness to give up certain claims about the meaning of expressions easier than others. Both meanings in DAT remain in principle revisable, and whether or not such a revision leads to a retrospective language change, attributing the new meaning to speakers prior to the change, may hinge on various linguistic and social factors. The resulting post-Quinean revisable and fallible notion of analyticity takes the edge off the Paradox of Analysis and related worries. Although it may have been part of the noumenal and core meanings of 'atom' that they cannot be split, making 'Atoms are the smallest, indivisible building blocks of nature' analytic in this weak sense, we all know that this definition eventually went up in flames.

There is a related critique following Moore (1903) based on a position that I wish to call 'semantic primitivism'. In this view a use of 'good' points to some primitive property of goodness, 'democracy' means democracy or the property of being a democratic state, 'electron' stands for the property of being an electron, and so forth. In support of their theory, primitivists resort to variants of the Paradox of Analysis, Moore's Open Question Argument, and sometimes an allegedly vicious definitory circle.[33]

I do not have much to say about these kind of arguments except that I find them fairly unconvincing. They contradict existing knowledge and are pragmatically incoherent with existing linguistic practices. We already believe that a democratic state is somehow defined by some of its characteristics, and we already believe that a 'good knife' must have certain features and lack certain misfeatures. A primitive, non-decomposable concept theory also does not tell us anything useful about meaning and gives us no information about how to falsify statements containing the respective value expressions. That being said, it is worth emphasizing that DAT is compatible with the thesis that (a) certain referential expressions like proper names have no linguistically mandated core meaning—a speaker may associate a meaning with them, but need not do so by virtue of linguistic competence—, and the thesis that (b) the noumenal meaning of certain expressions such as natural kind terms is fallibly extracted from our investigations of the entities and ultimately fixed indexically. However, I hope to have made it plausible that value terms and predicates to not generally work this way.

---

[33] Moral intuitionists like Dancy (2004) seem to find the circle argument appealing. Since you have to justify the conceptual analysis of the conceptual analysis, and so forth, at some point moral intuitions have to kick in. I find the idea that you have to justify everything at any arbitrary level of semantic decomposition and the idea that intuitions could serve as a justification equally unappealing, and in any case the same allegedly vicious circle can be applied to any lexical analysis of any expression, not just moral terms. The bottom line of my reply is that from a Platonic point of view the circle is not vicious and there is no need to start all of your justifications with the Big Bang. Justification is a dialectic process that stops with agreement or in an aporetic stalemate.

6. *Summary and Conclusions*

Competent speakers associate a core meaning with value terms, and it has been argued that the commendatory nature of value terms may be the best candidate for this common denominator among speakers. Speakers also generally strife to match, upon further reflection, some fitting aspect of reality with the use of such terms. Even for syncategorematic expressions like 'and' various noumenal meanings have been suggested in the logical and semantic literature, and perhaps only few or none of them fully match the core meaning that may have evolved from our need to coordinate behavior by communicating with each other. When even logical expressions have controversial noumenal meanings, then it ought not come as a surprise that there are many different candidates for the noumenal meaning of value terms. Perhaps we tend to not be satisfied with the core meanings of value terms despite pervasive disagreement about suggestions for their noumenal meaning, because that is generally the way language works. Once we transcend our needs of communicating for some immediate needs, become curious about aspects of reality that go beyond a mere need of coordinating with others, and want to describe reality correctly, we naturally assume that there is something in reality, the noumenon, to which the use of a term corresponds—which does, of course, not imply that there really is such a thing. On the basis of this thesis I submit that dialogues like (1)–(8) do not exemplify mere negotiations but rather attempts to balance noumenal meaning and core meaning in a way that approximates the noumenal meaning to a correct description of the often ephemeral and purported noumenon. If the noumenal definition deviates too much from the core meaning, then we obtain a technical definition that may appear to be arbitrary. If on the other hand the noumenal definition deviates too much from the noumenon, as for instance new evidence is acquired, then it is revised, which may in turn trigger a slower revision of the core meaning. So after all, disputes about value terms that are not directly content-based may not be so different from disputes about theoretical terms or terms for abstract objects for which an ostensive definition will not do either. They are peculiar because of the widespread disagreement about their noumenal meaning, but this may be the result of their often rather complex multi-dimensional comparative structure and the complicated psychological and social phenomena with which they are connected. This approach is compatible with the metalinguistic negotiation thesis and should be understood as a précis and extension of it.

References

Bach, K. 2005. "Context ex machina." In Z. G. Szabó (ed.). *Semantics versus Pragmatics*. Oxford: Oxford University Press: 16–44.

Burge, T. 1979. "Individualism and the Mental." *Midwest Studies in Philosophy* 4:73–121.

Burge, T. 1986. "Individualism and psychology." *Philosophical Review* 95: 345.

Burgess, A. 2013. "Conceptual ethics i." *Philosophy Compass* 8 (12): 1091–1101.

Burgess, A. and D. Plunkett 2013. "Conceptual ethics ii." *Philosophy Compass* 8 (12): 1102–1110.

Cappelen, H. and E. Lepore 2004. *Insensitive Semantics*. Oxford: Blackwell.

Chalmers, D. 2011. "Verbal disputes." *Philosophical Review* 120 (4): 515–566.

Cresswell, M. J. 1990. *Entities and Indices*. Dordrecht: Kluwer.

Cresswell, M. J. 1996. *Semantic Indexicality*. Dordrecht: Kluwer.

Dancy, J. 2004. *Ethics without Principles*. Oxford: Clarendon Press.

de Sá, D. L. 2008. "Presuppositions of commonality: An indexical relativist account of disagreement." In M. Garcia-Carpintero and M. Kölbel (eds.). *Relative Truth*. New York: Oxford University Press.

de Sá, D. L. 2009. "Relativizing utterance-truth?" *Synthese* 170 (1): 1–5.

de Sá, D. L. 2015. "Expressing disagreement." *Erkenntnis* 80: 153–165.

Egan, A. 2010. "Disputing about taste." In R. Feldman and T. Warfield (eds.). *Disagreement*. New York: Oxford University Press: 247–286.

Egan, A. 2014. "There's something funny about comedy: A case study in faultless disagreement." *Erkenntnis* 79: 73–100.

Gärdenfors, P. 2000. *Conceptual Spaces*. Cambridge: MIT Press.

Hare, R. 1952. *The Language of Morals*. Oxford: Oxford University Press.

Hare, R. M. 1999. *The Language of Morals*. Oxford: Clarendon Press.

Horgan, T. and M. Timmons 1992. "Troubles on moral twin earth: Moral queerness revived." *Synthese* 92: 221–260.

Kaplan, D. 1989. "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals." In J. Almog, J. Perry, and H. Wettstein (eds.). *Themes from Kaplan*. Oxford, New York: Oxford University Press: 481–564.

Kölbel, M. 2002. *Truth without Objectivity*. London: Routledge.

Kripke, S. A. 1972. "Naming and Necessity." In G. Harman and D. Davidson (eds.). *Semantics of Natural Language*. Dordrecht, Boston: D. Reidel Publishing Co: 253–355.

Lasersohn, P. 2005. "Context dependence, disagreement, and predicates of personal taste." *Linguistics and Philosophy* 28 (6): 643–686.

Lewis, D. K. 1969. *Convention: A Philosophical Study*. Harvard: Harvard University Press.

Lewis, D. K. 1980. "Index, Context and Content." In S. Kanger and S. Öhman (eds.). *Philosophy and Grammar*. Dordrecht, Boston: D. Reidel Publishing Co: 79–100.

Ludlow, P. 2008. Cheap contextualism. *Philosophical Issues* 18: 104–129.

MacFarlane, J. 2005. "Making sense of relative truth." *Proceedings of the Aristotelian Society* 105: 321–39.

MacFarlane, J. 2012. "Relativism." In D. G. Fara and G. Russell (eds.). *The Routledge Companion to the Philosophy of Language*. New York: Routledge: 132–142

MacFarlane, J. 2014. *Assessment Relativity: Relative Truth and its Applications*. Oxford: Clarendon Press.

Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. London: Penguin.

Marques, Teresa and García-Carpintero, M. (2014). "Disagreement about Taste: Commonality Presuppositions and Coordination." *Australasian Journal of Philosophy* 92 (4): 701–723.

Meinong, A. 1971. "Über Gegenstandstheorie." In R. Haller (ed.). *Alexius Meinong Gesamtausgabe: Abhandlungen zur Erkenntnistheorie und Gegenstandstheorie*, Volume II. Graz, Austria: Akademische Druckund Verlagsanstalt Graz: 481–535.

Moore, G. E. 1903. *Principia Ethica*. Cambridge: Cambridge University Press.

Penco, C. 1999. "Objective and cognitive context." In M. Benerecetti and F. Castellani (eds.). *Modeling and using context : Lecture Notes in Artificial Intelligence 1688*. Springer: 270–283.

Plunkett, D. and T. Sundell 2013. "Disagreement and the semantics of normative and evaluative terms." *Philosophers' Imprint* 13 (23): 1–37.

Plunkett, D. and T. Sundell 2014. "Antipositivist arguments from legal thought and talk: The metalinguistic response." In G. Hubb and D. Lind (eds.). *Pragmatism, Law, and Language*. London: Routledge: 56–75.

Putnam, H. 1975a. "Is semantics possible?" In H. Putnam (ed.). *Mind, Language and Reality*. Cambridge: Cambridge University Press: 187–201.

Putnam, H. 1975b. "The Meaning of 'Meaning'." In H. Putnam (ed.). *Mind, Language and Reality*. Cambridge, UK: Cambridge University Press: 215–271.

Putnam, H. 1994. "The meaning of 'meaning'." In R. M. Harnish (ed.). *Basic Topics in the Philosophy of Language*. New York, London, Tokyo: Harvester Wheatsheaf: 221–274.

Putnam, H. 2002. *The Collapse of the Fact/Value Dichotomy and Other Essays*. Cambridge: Harvard University Press.

Quine, W. V. O. 1964. *From a Logical Point of View*. Second revised edition. Oxford: Oxford University Press.

Rast, E. H. 2009. "Context and interpretation." In J. M. Larrazabal and L. Zubeldia (eds.). *Meaning, Content, and Argument. Proceedings of the ILCLI International Workshop on Semantics, Pragmatics, and Rhetoric*. San Sebastian: University of the Basque Country Press: 515–534.

Rast, E. H. 2011. "Classical possibilism and fictional objects." In F. Lihoreau (ed.). *Fiction in Philosophy*. Frankfurt: Ontos.

Rast, E. H. 2014. "Context as assumptions." In F. Lihoreau and M. Rebuschi (eds.). *Knowledge, Context, Formalism*. Springer: 9–39.

Recanati, F. 2004. *Literal Meaning*. Cambridge: Cambridge University Press.

Recanati, F. 2006. Crazy Minimalism. *Mind and Language* 21 (1): 21–30.

Rosch, E. 1975. "Cognitive representations of semantic categories." *Journal of Experimental Psychology: General* 104 (3): 192–233.

Rosch, E. 1983. "Prototype classification and logical classification: The two systems." In E. Scholnick (ed.). *New Trends in Conceptual Representation: Challenges to Piagets Theory?* Hillsdale: Lawrence Erlbaum Associates: 73–86.

Stojanovic, I. 2015. "Evaluative adjectives and evaluative uses of ordinary adjectives." In D. Bekki and E. McCready (eds.). *Proceedings of LENLS12: Language Engineering and Natural Language Semantics*. The Japan Society for Artificial Intelligence.

Sundell, T. 2011. "Disagreements about taste." *Philosophical Studies* 155 (2): 267–288.

Väyrynen, P. 2013. *The Lewd, the Rude and the Nasty: A Study of Thick Concepts in Ethics*. New York: Oxford University Press.

von Wright, G. H. 1963. *The Varieties of Goodness*. London: Routledge & K. Paul.

Wittgenstein, L. 1969. *Tractatus logico-philosophicus*. London: Routledge.

Zalta, E. 1983. *Abstract Objects*. Dordrecht, Boston, Lancaster: D. Reidel.

Temporally Restricted Composition

MARK STEEN
Allegheny College, Meadville, USA

I develop and defend a novel answer to Peter van Inwagen's 'Special Composition Question,' (SCQ) namely, under what conditions do some things compose and object? My answer is that things will compose an object when and only when they exist simultaneously relative to a reference frame (I call this 'Temporally Restricted Composition' or TREC). I then show how this view wards off objections given to 'Unrestricted Mereology' (UM). TREC, unlike other theories of Restricted Composition, does not fall prey to worries about vagueness, anthropocentrism, or arbitrariness. TREC also has advantages over all the other answers to the SCQ. TREC is an account an A-theorist anti-Eternalist who wants an unrestricted mereology should accept. I also engage in some conceptual hygiene by showing how UM, as it should be used, should not, in itself, entail or contain a commitment to either Eternalism or Four-Dimensionalism.

Keywords: Mereology, special composition question, part-whole relation, naturalized metaphyics, composition, special relativity theory.

Peter Van Inwagen's 'Special Composition Question' ('SCQ') asks—when do some things compose or make up a further thing? (Van Inwagen 1990: 21–33). The Unrestricted Mereologist answers—'always'. (see Rea 2008 and Sider 2001) Some rivals of UM, the Nihilists, answer: 'never' (e.g. Rosen and Dorr 2002, Dorr 2005). Organicists say composition occurs 'when the things compose an organism'. (Van Inwagen 1990, Merricks 2003) According to the folk theory we could call 'Intuitivism' the answer is that composition occurs 'whenever we intuitively think it does.'[1] Lastly, the Brutalist answers that composition's obtaining or not is brute and unexplainable (Markosian 1998).

There are serious problems with all of these accounts. No one, to my knowledge, has proposed the answer that I will recommend.[2] I will

---

[1] This view is tacitly assumed by many, e.g. Wiggins 1980.

[2] There is however, brief discussion of one of the embedded conditionals in Hudson 2001. As this is nearing publication, I just noticed the paper by McKenzie

argue that the correct answer to the SCQ is what I will call 'Temporally Restricted Composition', or 'TREC' for short. Stated informally it is as follows:

(TREC) For any X's they will compose a fusion F at a time t when and only when all of their parts exist simultaneously, relative to a reference frame.[3]

Stated informally without regard to Relativity Theory, we have:

(TREC*) For any X's they will compose a fusion F when and only when the X's exist simultaneously.

I will argue that TREC is a better answer to the SCQ than the rival positions. TREC acknowledges all the composite objects of common-sense, respects many of the persuasive arguments which support un-restricted mereology, without incurring ontological commitment to a profusion of cross-time fusions. TREC is a non-*ad hoc* answer to SCQ which avoids arbitrariness and anthropocentricism, while also avoid-ing problems of vagueness.

I will suppose that the fundamental objects/things over which the SCQ ranges are either fundamental or simple particles (if there are any) or 'atomless gunk' portions (if there are any). Events, if there are any, would I believe have a different mereological structure than ob-jects, and when people talk of objects they are not talking of baseball games or recessions. At least, this seems presumed in the literature on composition, which has sidestepped events (with the exception of 4d, which one could construe as consolidating the categories of *object* and *event*). Certainly the literature is largely restricting the discussion with SCQ to objects, most likely to its detriment. And it also seems that TREC is decidedly *not* true of events, such as baseball games which are composed of nine non-simultaneous innings. So, none of what I say about objects will necessarily hold of events.

TREC is inconsistent with Four Dimensionalist accounts which ac-cept that there are wholes composed of temporally disjoint parts. TREC is only consistent with Three-Dimensionalism, which is the doctrine that objects persist through time by being 'wholly present' at different times, rather than by having distinct temporal parts at different times.[4] One of my targets is Unrestricted Mereology, but this is not equivalent

and Muller 2017, which has some similarities, but based on state boundedness rather than simultaneity.

[3] Slightly more formally, $\forall t \forall F \forall X : (\exists y : XCyt <=> XEtF)$, or, for all time's t, frames of reference F, and any X's, there will exist a y such that the X's compose y at t if and only if the X's Exist-at-t-relative to F. Thanks to two anonymous referees for some helpful comments on TREC.

[4] TREC does not strictly speaking entail Three-Dimensionalism, since it is consistent with the claim that simples never last more than an instant, according to any reference frame. TREC is only Three-Dimensionalist along with this supposition, which is a way of assuming Three Dimensionalism. I will suppose that simples can, and often do, endure.

to Four-Dimensionalism. Unrestricted Mereology entails that temporal parts compose cross-time fusions only if temporal parts exist. One could furthermore be a Four-Dimensionalist who has a restriction on which temporal parts fuse. I will just assume Three-Dimensionalism, and show how TREC gels with a Three-Dimensionalistic picture.

Intuitively, we want there to be a restriction on composition, otherwise we will have strange objects, composed, for instance, of Alexander the Great's kidney and Tom Cruise's gall bladder, and another composed of all the salmon and Obama's great-great grandchildren. But, the main arguments against a restriction on composition is that any proposed restriction will be either (i) arbitrary or *ad hoc*, (ii) anthropocentric, or (iii) will entail worldly vagueness about what exists (Sider 2001: 11–74). I will show that TREC succumbs to none of these objections. Then I will show what benefits TREC has, compared to its rivals.

TREC is not arbitrary. Simultaneity according to a reference frame is a significant natural property, which arguably must obtain for any causal relations to obtain.[5] It is not picked out of the blue to obtain certain desired results. It is a non-Presentist acknowledgment of simultaneity as crucial in one's consideration for objects to 'get together' to make up a further object. Objects which are not on the same hyperplane, and are outside of each other's future and past light cones can have no causal interactions. It is intuitive that a composite object's parts can exert some causal influence on each other, and act jointly (in such a way, for instance, that they can have a center of gravity, which cross-time fusions cannot). Supposed fusions which can do none of these things seem to lack what it takes to compose an object. However, objects which do exist simultaneously according to a reference frame, such as all those which compose our galaxy, *can* have causal interactions, such as mutual gravitational pull, shared spin-relations along an axis or center of gravity, and so on.

Three-Dimensionalists should be suspicious of cross-time fusions. Assume that there is a world W with absolute time, and suppose that x exists at t1 only, at y at t3 only. The unrestricted mereologist would hold that x and y compose a fusion F. But the Three-Dimensionalist, especially of the A-theorist variety, should doubt this. When does F exist? Not at t2. An object can't exist at a time when none of its parts do. At t1? How could *it* (the fusion of x and y) exist at t1? Only part of it exists then. Y is not even in existence yet, so how could it compose something? At t3? How could it be? One of its parts, x, no longer exists? I do not take this as conclusive, and to solidify this point would take us far afield of the goals of this article. I merely want to point out

[5] Quantum entanglement phenomena does not cut against this, as entangled particles needed to exist simultaneously at some time according to a reference frame in order to be entangled, even if the effects of this entanglement could manifest after the particles are no longer simultaneous according to any reference frame. C.f. McKenzie and Muller 2017.

how the Three-Dimensionalist Non-Presentist[6], who already believes in some kind of temporally relativized notion of property-possession (such as shapes, which are often held to be disguised relations to times by Three-Dimensionalists), has ample reason to believe that parthood is temporally relative and has a synchronous requisite as well, if the 'present' in 'wholly present' is to have any meaning. TREC does not itself entail mereological essentialism (ME), or the doctrine that objects have all of their parts essentially, unless we assume that all objects are identical with mereological sums, which is questionable. (More on this below). But TREC it is consistent with ME, and can be combined with standard mereological theses such as uniqueness and extensionality which will entail ME. I do think that ME is correct, and while TREC helps make sense of and defend it, the truth of TREC does not depend on it.[7]

TREC is not anthropocentric. Even those who believe that time does not 'really' exist, or believe that the flow of time is illusory, still accept, if they accept Relativity Theory, that simultaneity according to a reference frame, light-cone charts, and Minkowski spacetime diagrams, all capture important natural joints in the world which are neither psychological nor merely conventional.

While some still believe that reference frames are necessarily defined in terms of an observer, most do not, and can define a reference frame in terms of a coordinate system. Most believe that Relativity phenomenon still occur where no observers are, or could be. If this is so, then TREC does not succumb to the anthropocentricism charge on this score, either.

TREC does not entail worldly vagueness. One of the strongest arguments against restrictions on composition has been given by Ted Sider (2001 chapter 4.9), and now goes by the name of 'the argument from vagueness,' which is used to support Four Dimensionalism. Briefly, and superficially (the argument is quite complicated), the problem, supposedly, with any restriction on composition short of Nihilism (the doctrine that composition never occurs) is that the restriction will inevitably be vague, so that it is indeterminate what fusions there are, and hence indeterminate how many things exist, which is impossible.

Take any restriction you like (e.g., falling under a commonsense sortal, being chemically bonded, composing a life, etc.). Sider states that for any candidate restriction R captured by conditions C we can create a sorites-like series of cases starting from C determinately holding to a situation where it is indeterminate whether C obtains, and hence whether the fusion in question exists, and hence whether an object exists, which is impossible (Sider 2001 Ch. 4.9).

---

[6] I take it that Presentism is not compatible with Relativity Theory, and hence a Presentist can't accept TREC*, which denies objective simultaneity. Although, see Zimmerman 2011 about how Presentism and RT could be reconciled.

[7] See Steen (2016) for a defense of mereological essentialism.

TREC does not succumb to this objection if synchrony is not conventional. While this issue is somewhat controversial, it is held by many that Malament's Theorem proves the non-conventionality of simultaneity according to a reference frame. Torretti says,

> Malament proved that simultaneity by standard synchronism in an inertial frame F is the *only* non-universal equivalence between events at different points of F that is definable ("in any sense of 'definable' no matter how weak") in terms of causal connectibility alone, for a given F. (Torretti 1983: 229)

So, if synchrony is objective, then the argument from vagueness has no bite against the TREC theorist. There is no sorites-series of time indexed states of affairs S1...Sn, where, according to S1, it is determinate that composition holds in S1, and indeterminate whether composition holds for some first S1+n. For any two objects O1 and O2, and for any arbitrarily chosen reference frame R, either O1 and O2 exist simultaneously according to R or not. If there is no R such that O1 and O2 exist simultaneously according to it, then there will be no fusion of them whatsoever.

If one objects that a fusion F composed out of O1 and O2 will exist according to a frame R1, and not according to R2, and hence it is indeterminate whether F exists, then one is not thoroughly accepting Relativity Theory. One would be, against Relativity Theory, privileging one frame over another, or, incoherently, attempting to combine a timeless notion of parthood with the TREC view that parthood is relative to times (which are in turn relative to reference frames). In fact, according to TREC, parthood is a four-place relation. There is no parthood simpliciter. O1 cannot simply be a part of O2. O1 can only be a part of O2 if O1 is a part of O2 at time t, relative to R.

It might be thought strange that there is no frame-independent answer to how many things there are, but this amounts to no more than an incredulous stare. There are also no frame-independent answers to how long something is, when something occurred, how much time passed, and so on. There are also no frame-independent temporal orderings of a sequence of events. E1 can occur before E2 according to R1, whereas E2 can occur simultaneous with E1 according to R2. If one accepts Relativity Theory already, with its concomitant strangeness, then the relativity of which composites there are is par for the course.

Lastly, some parting words on why TREC is superior to its rivals. TREC has all of the strengths of UM, but lessens the bite of its main weakness. UM is supported by arguing against arbitrariness, anthropocentricism, and vagueness in composition. TREC endorses all of these arguments, and obeys the same strictures, without entailing a bevy of cross-time fusions which can exist even when some or none of their parts do (or exist eternally as parts of one thing, even when there is no time at which it exists). TREC also blocks the argument from vagueness, and gels with Three-Dimensionalism, which has support

independent of worries about composition. (There will be more on the objection of how the motivation for TREC really serves as a motivation for UM below).

TREC is superior to 'Organicism', or the view that the only composites are living beings. This is a severely limited view on what composites there are. TREC allows for living beings, as well as buildings, tables, and rocks. Organicism can be viewed also as 'inanimate composite object nihilism,' and has severe weaknesses in that 'causal-redundancy' and compositional scepticism premises command less credence than the proposition they entail the falsity of, namely that there are composite non-living objects. Organicism also has the unwelcome consequence of either positing a sharp-line between the living and the non-living, which doesn't seem to exist in nature, or supporting that there are some vague objects (namely, objects of which it is indeterminate whether they are living, and hence indeterminate whether there are certain sums).

TREC is superior to Nihilism, or the view that there are no composite entities whatsoever. TREC allows composite objects, which is obviously an intuitive desiderata, and is untouched by one of the major concerns that motivate Nihilism, namely, concerns about vagueness. Also, the Nihilist who rejects extended simples cannot hold that there is 'gunk', or stuff whose proper parts all have proper parts, *ad infinitum*. There can be no gunk, according to the Nihilist who rejects extended simples, since gunk is composite, and the gunk hypothesis rules out point particles.[8] TREC has no such entailments, and does not rule out gunk worlds.

TREC is superior to 'Brutal Composition,' or the view that whether or not composition occurs is just a brute fact with no explanation whatsoever (Markosian 1998). Any well-supported explanation is better than none, and TREC gives one. Furthermore, this explanation goes along with commonsense more than Brutalism, in that Brutalism leaves it mysterious whether there are ordinary composite objects, and also allows, in principle (since there is no principle disallowing it) that there are bizarre cross-time fusions. The Brutalist rules out Four-Dimensionalism via the argument from vagueness only by positing a brute restriction which, for all we know, rules out all the composite Three Dimensional objects the Brutalist was attempting to rescue.

TREC is also superior to the often held, but almost never stated view of composition which we could call 'Intuitivism'. Intuitivism is the view that composition occurs whenever we intuitively think it does. Intuitivism is also the view held by those who think worrying about the Special Composition Question is silly. Such people, rather than thinking there is no answer to the SCQ, often believe that there are some

---

[8] And, the acceptance of extended simples brings along it's own problems. Furthermore, the supposition of there being stuff which is not things which composes simples goes against the desired sparseness of the Nihilists view. For work on extended simples, see Markosian 1998a and 2004.

composites (such as themselves, or baseballs) but not others (such as the object composed of the Pyramids of Giza and the Bee Gees), but either think that there is no principle whatsoever, or the unsatisfactory Intuitivist one. This 'view', if it can be called one, also succumbs to the charges of arbitrariness, anthropocentricity, and vagueness. One cannot escape the SCQ by refusing to think about it, since, by default, one will accept this unacceptable folk view about composition. One might attempt to rescue Intuitivism by becoming Aristotelian or neo-Aristotelian, and hold that simples compose a sum S just in case there is an Aristotelian form F (or substantial form, or rigid embodiment, or dominant sortal, etc. etc.)[9] which inheres in the simples. But, if one does this, one just reintroduces the worries about anthropocentricity and vagueness. Which candidate F's are actually forms? Can one rule out bizarre substantial forms (or whatever substitute you like) and give an answer to the SCQ without being slavish to our contingently possessed perceptual and cognitive faculties, and interests? Can one hold a view like this and rule out vagueness? I hold, with Sider (2001) and Kurtsal Steen (2010), that the answer is 'no'.

A final worry about TREC is that, while it does rule out all cross-time fusions, that it does allow in many strange ones not countenanced by the folk, such as the sum of the paper you are reading and your pinky. But, I think the argument from vagueness, while not showing that *any* restriction is nonviable (for I believe of course that TREC is a viable restriction), does in fact show that every other presented alternative to Unrestricted Mereology and TREC is nonviable. While there may be other restrictions which are neither arbitrary, anthropocentric, or entail worldly vagueness, I have not seen any. Kurtsal Steen (2010) does a good job of showing how, despite appearances, the argument from vagueness does not in fact entail four-dimensionalism. So, my employment of it in arguing against other views and for TREC does not commit me to four-dimensionalism either.

But one might think that the TREC is not sufficiently motivated since vagueness, arbitrariness, and anthropocentricity concerns support UM. TREC still allows in a bevy of odd objects, and faces the same objections that UM does. So why not 'go all the way' for wholesale UM?

There are several reasons not to. One should ask the following—Does UM entail Eternalism (the doctrine that all times, past present and future are all equally real, there being no privileged 'now')? If so, then all those reasons to reject Eternalism (e.g free will issues, the problem of change, counterintuitiveness, no time flow, etc. etc.) apply to UM. If not, then UM may in fact be consistent with TREC (e.g., if Presentism were true, a UM and TREC would entail the existence of a coextensive sets of fusions).

But, if one looks at the literature, and how UM is understood, its proponents do in fact seem to believe that UM entails Eternalism or

---

[9] See, for example, Burke 1994, Fine 1999, and Koslicki 2008.

has Eternalism as a part of the doctrine.[10] But, as I think I have shown, one can think that whatever exists has a fusion and yet reject UM, and so there is a major distinction between UM'ists and TRECkies. TREC is like unrestricted mereology for A-theorists. To the extent that one accepts A-theory and the arguments for restricted composition based on the worries of arbitrariness etc. then one should accept TREC.

There are other arguments one could give for TREC over UM, such as some of the table-pounding one given on page 3. One might also say that UM falls prey to 'double-counting' in a way that TREC does not. Suppose that there are a collection of simples spread around the galaxy at t1—call it S1. Now suppose, as a three-dimensionalist would say, that the very same collection of simples endures (no parts go out of existence) and is identical with a collection S2 at t2. This seems innocent (to me). The TRECkie would say that S2 is S1 and we have one entity. The UM'ist would say that S2 and S1, being distinct entities, would compose a fusion S3, which exists atemporally. But this strikes me as double-counting.

Or would they accept S3 as a distinct entity? Well, this depends—does UM entail Four-Dimensionalism, or have Four-Dimensionalism as a part of the doctrine? If it does, then those who do not accept Eternalism and Temporal Parts would have reason to reject it, and would need to posit some other principle, similar to UM, which does not entail these views. (And what would be if not TREC?). If it does not entail 4d, then it seems that UM is actually consistent with TREC. But, the way the phrase is used, it does in fact seem that UM is a 'package deal', namely, unrestricted mereology + Eternalism + Four-dimensionalism (of the temporal parts variety).

All I have done is merely outline and argue for a modest proposal—namely, a more narrow theory of unrestricted composition for A-theorist Three-Dimensionalists who accept that restrictions on composition are hopeless. By pointing out that the phrase 'unrestricted mereology' or 'unrestricted composition' are misleading, I've teased apart several features that get run together and laid out how a Three-Dimensionalist A-theorist who is moved by the argument from vagueness should think about unrestricted composition. Hopefully that might clear up some problems in the debates about persistence and time. At the very least, I have laid out what a theory of unrestricted composition might look like if one accepts, say, the Moving Spotlight, or the Growing Block views of time.[11]

References

Benovsky, J. 2016. "Eliminativism and gunk." *Teorema: International Journal of Philosophy* 35 (1): 59–66.

Burke, M. B. 1994. "Preserving the Principle of One Object to a Place: A Novel Account of the Relations Among Objects, Sorts, Sortals, and Persistence Conditions." *Philosophy and Phenomenological Research* 54 (3): 591–624.

Cameron, R. 2015. *The Moving Spotlight: An Essay on Time and Ontology*. Oxford: Oxford University Press.

Chisholm, R. 1976. *Person and Object*. La Salle: Open Court Publishing.

_____ 1986. *Roderick M. Chisholm: Self-Profile*. Bogdan, R.J. (ed.), Dordrecht: D. Reidel.

Dorr, C. 2005. "What we disagree about when we disagree about ontology." In M. Kalderon, (ed.). *Fictionalism in Metaphysics*. Oxford: Oxford University Press: 234–286.

Fine, K. 1999. "Things and Their Parts." *Midwest Studies in Philosophy* 23 (1): 61–74.

Grandy, R. E. 1975. "Stuff and Things." *Synthese* 31: 479–485.

Hudson, H. 2001. *A Materialist Metaphysic of the Human Person*. Ithaca: Cornell University Press.

Janis, A. 2006. "Conventionality and Simultaneity." *Stanford Encyclopedia of Philosophy*. < URL = http://plato.stanford.edu/entries/spacetime-convensimul/

Karmo, T. 1977. "Disturbances." *Analysis* 37: 147–148.

Koslicki, K. 2008. *The Structure of Objects*. Oxford: Oxford University Press.

Kurtsal Steen, I. 2010. "Three-dimensionalists Semantic Solution to Diachronic Vagueness." *Philosophical Studies* 150: 79–96.

Le Bihan, B. 2013. "Why a Gunk World is Compatible with Nihilism about Objects." *Studia Philosophica Estonica* 6 (1): 1–14.

Malament, D. 1977. "Causal Theories of Time and the Conventionality of Simultaneity." *Noûs* 11: 293–300.

Markosian, N. 1998a. "Brutal Composition." *Philosophical Studies* 92 (3): 211–249.

_____ 1998b. "Simples." *Australasian Journal of Philosophy* 76 (2): 213–228.

_____ 2004. "Simples, Stuff and Simple People." *The Monist* 87: 405–428.

McKenzie, M. K., and Muller, F. A. 2017. "Bound States and the Special Composition Question," In Massimi M., Romeijn JW., Schurz G. (eds.).

EPSA15 Selected Papers. European Studies in Philosophy of Science vol. 5. Springer

Merricks, T. 2003. *Objects and Persons*. Oxford: Oxford University Press.

Quine, Willard (1976). Whither physical objects? In Cohen, R. S., Feyerabend P. K. and Wartofsky M. (eds.). *Essays in Memory of Imre Lakatos*. Reidel: 497–504.

Rea, M., 1998. "In Defense of Mereological Universalism." *Philosophy and Phenomenological Research* 58: 347–360.

Rosen G., and Dorr C. 2002. "Composition as fiction." In Gale R (ed.). *The Blackwell guide to metaphysics*. Malden: Blackwell: 151–174

Sider, T. 2001 *Four-Dimensionalism. An Ontology of Persistence and Time.* New York: Oxford University Press.

Steen, M. 2016. "The Metaphysics of Mass Expressions." *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/metaphysics-massexpress/>.

Torretti, R. 1983. *Relativity and Geometry*. Oxford, New York: Pergamon.

Van Inwagen, P. 1990. *Material Beings.* Ithaca: Cornell University Press.

_____ 2009. "Can Mereological Sums Change Their Parts?" *Journal of Philosophy* 103 (12): 614–630.

Wiggins, D. 1980 *Sameness and Substance.* Cambridge: Harvard University Press.

Zimmerman, D. 1995. "Theories of Masses and Problems of Constitution." *Philosophical Review* 104 (1): 53–110.

_____ 2011, "Presentism and the Space-Time Manifold." In Callender C. (ed.). *The Oxford Handbook of Time.* Oxford: Oxford University Press: 163–244.

Book Discussion

The Power of Language: Discussion of Charles Taylor's The Language Animal: The Full Shape of the Human Linguistic Capacity

JOŠKO ŽANIĆ
Zagreb

The paper is a discussion of Charles Taylor's recent book The Language Animal. *The criticism of Taylor's view of language clusters around two main themes: first, that he seems to "mysterianize" language somewhat, whereas the topics he addresses can be adequately dealt with within standard formal approaches in the philosophy of language and cognitive science; second, that his focus on language is in many cases misplaced, and should indeed be replaced with a focus on human conceptual structure, which language only fragmentarily expresses.*

Keywords: Human linguistic capacity, human meanings, value.

Charles Taylor's most recent book[1] contains, towards the end, this sentence which indicates its scope and ambition: "With phrases like 'animal possessing language', we are trying to answer a question like: 'what is human nature?'" (338–9). This aligns very closely with the leading thought of a book published also in 2016 by another crucial contemporary thinker, Noam Chomsky, who states: "there are much more fundamental reasons to try to determine clearly what language is, reasons that bear directly on the question of what kind of creatures we are" (2016: 2). But, if their goals in inspecting language are closely related, the respective takes on language by Chomsky and Taylor couldn't be more different. Whereas Chomsky is one of the founders of contemporary formal approaches to language, Taylor is concerned

---

[1] See references for bibliographical data.

throughout his book with stressing the affective, enactive, "embodied" aspects of language not captured by the formal theories; and whereas Chomsky is the arch-internalist and individualist of our time, Taylor insists on seeing language in the context of sharing ("the linguistic capacity is essentially shared" 333) and "communion" (situations of joint attention).

So, what is Taylor's exact take on language in the book that is the focus of this discussion? He states: "language can only be understood if we understand its constitutive role in human life" (261). Language brings to human life new goals and purposes that wouldn't exist without it, it alters our way of being in the world (echoes of Heidegger are non-accidental)—language is not just part of the framework of general human activities, but transforms the whole framework. Language is crucially *constitutive*, claims Taylor drawing on the work of Herder, in opposition to being purely designative of antecedently given entities, a position he identifies in the work of Hobbes, Locke and Condillac (what he calls the HLC theory), and continued in "post-Fregean analytic philosophy of language".

Crucial aspects of the linguistic capacity which demonstrate this constitutivity for Taylor are three: what he calls human or metabiological meanings; the efficacy of discourse; the sense-making power of narrative. So, let me explicate them briefly.

As examples of human meanings, Taylor lists virtue and motivational terms (generosity, loyalty, love/lust), terms for stances and ways in which we experience life (serene, troubled) and aesthetic terms. These meanings, as opposed to mere "biological meanings" (like "food", for instance), cannot be understood from the outside and dispassionately; instead, they are *felt*, they cannot exist without affect, and require *Einfühlung* (empathizing, putting oneself in the other's shoes) to be understood. Further, these meanings, claims Taylor, are interconnected, forming "skeins", and are deeply interwoven with culture (think of the notion of loyalty for a samurai and a member of contemporary individualistic society).

Crucially for Taylor, human meanings open up new domains, new ways of experiencing. They are not just names for antecedently existing and language-independently observable things. Rather, the expression used for the meaning is essential to the experience it names, not preceded by it. These meanings are in such a way constitutive of specifically human ways of being. Further, although they are dependent on us, we can, claims Taylor, get things wrong when operating with these meanings—both descriptively (wrongly describing somebody as loyal) and normatively (wrongly attacking somebody for lack of loyalty). And we can correct ourselves, through "transitions" (development of better understanding) ratified by intuition.

These meanings crucially involve enactment, claims Taylor—for instance, the display of loyalty involves types of bodily movement and

bearing. They can also emerge through works of art. These works, in case of literary works of art, contain no assertions about life and human meanings, but can *portray* them. In fact, Taylor identifies three levels of articulation of meanings (three "rungs in a ladder", as he puts it): enactive; symbolic (which concerns the portrayals that works of art offer); and descriptive-analytic.

Human meanings are, according to Taylor, understood hermeneutically: in circles, without a final and objective definition.

The second aspect of constitutivity of language that Taylor discusses is the efficacy of discourse. Norms, footings, institutions and social orders are constituted and transformed in discourse, claims Taylor. By "footings" he understands the mutual positions of interlocutors in social-discursive space. These are affirmed in discourse (talking politely to one's superior), but can be challenged and transformed ("go to hell, boss"). Whole social orders are carried and shaped in discourse.

The third aspect is narrative. Narrative, says Taylor, offers insight through diachronic gestalts, units of character, event, motivation, etc., where the issues involved concern human meanings. Taylor claims that narrative is crucial to self-understanding (a claim well known from a strand of contemporary research into the constitution of personality). The telling of stories is therefore a creative/constitutive feature of language.

Before discussing these crucial aspects, Taylor discusses figuring, or the creative use of language. In "figuring A through B", mainly in metaphor, we gain new powers of articulation, with bodily know-how underlying this (here Taylor draws heavily on the work of the cognitive linguist George Lakoff and collaborator Mark Johnson who stress the creative role of metaphor and the bodily basis of metaphorical meanings). This is something that the HLC paradigm, according to Taylor, completely misses. Taylor concedes that Frege made some crucial adjustments to this paradigm: by introducing the context principle, he freed it of atomism; by construing sense as abstract and public, he cured it of individualism. However, this paradigm is, in Taylor's view, still limited by recognizing only designative (as opposed to constitutive) logic, which it attributes to language: it sees language as objective depictive power, and nothing above this. Taylor challenges this view throughout the book, as indicated above.

But before discussing language as constitutive, Taylor attributes to it other important characteristics. First of all, crucial to operating with language is an irreducible sensitivity to rightness (echoes of later Wittgenstein are obvious). Words have right and wrong uses, and an awareness of this is ineliminable as an aspect of the language capacity. Further, language is impossible unless holistically conceived—one word presupposes all others, does not function in isolation (Fodor would have a lot to say about this, but that will not be the focus of this discussion). Finally, language is part of a range of symbolic forms, which all have to

be taken into account if language is to be appreciated in its full scope.

Language (along with the self) develops, claims Taylor, following the work of Michael Tomasello, in the context of emotion-infused joint attention, which Taylor calls "communion". Of course, the ineliminable emotional chargedness of situations enables us to acquire the human meanings which are, as we have seen, "felt".

Language alters, Taylor claims, our way of being in the world, it opens up for us new dimensions of existence, which wouldn't be possible without it.

Criticism of Taylor's view of language will cluster around two main themes: first, that he seems to "mysterianize" language somewhat, whereas the topics he addresses can be adequately dealt with within standard formal approaches in the philosophy of language and cognitive science; second, that his focus on language is in many cases misplaced, and should indeed be replaced with a focus on human conceptual structure, which language only fragmentarily expresses.

Let me return to Taylor's "human meanings". First of all, Taylor seems to claim that many forms of thought would be utterly impossible without language—e.g. thought about loyalty. However, empirical research has shown this to be false. Steven Pinker's (1994), which Taylor cites, provides abundant evidence of this (cf. ch. 3). There are fully intelligent aphasics, capable of complex mental operations, such as playing card games or recounting pantomimed narratives. There are also languageless beings, such as prelinguistic infants or monkeys, who are able to reason about space, time, objects, number, causality and, interestingly with regard to the example of loyalty, obligations (in vervet monkeys) to members of family, such as avenging a member. Finally, many creative people report that their crucial insights didn't come through language, but through mental imagery.

On the other hand, if we take into account what kind of terms Taylor gives as examples of his human meanings, we see that these are simply value-laden terms, or thick concepts, in Bernard Williams's usage. And these, mostly in the guise of pejoratives or slurs, are actually a hot topic in contemporary philosophy of language. The most recent book-length contribution to the debate is Nenad Miščević's and Julija Perhat's (2016). This book offers a detailed, layered account of the meaning of such terms. For any pejorative (or laudative) it factors out its meaning into the following layers: minimal descriptive, descriptive-evaluative (e.g. "primitive, lazy, dangerous"), prescriptive (e.g. "to be avoided, discriminated against") and expressive (e.g. "yuck"). This account then offers a basis for claiming that using a pejorative or a laudative gives rise to several propositions being expressed by a sentence in which it is used. Now, this kind of account lends itself to realization in computational terms, and seems to take away a lot of the "aura" that Taylor builds around his human meanings—it places those meanings squarely within something that can be dealt with by a formal account of language. For example, the term "loyalty" could be cashed

out as having descriptive (the set of loyal acts), descriptive-evaluative ("proneness to help those one is obliged to"), prescriptive ("be such!") and expressive ("commendable!") layers.

As for the notion of human value itself, Ray Jackendoff, in his (2007), offers a whole theory of the human value system in computational terms. Jackendoff posits that the value system is a multidimensional calculating system, a part of the cognitive system which helps govern action. By multidimensional he means that different kinds of value can be distinguished, each with a valence and a magnitude, and with subjective and objective versions. The types of value Jackendoff identifies are affective value, utility, resource value, quality, prowess, normative value, personal normative value, and esteem. Important relations obtain between these kinds of value: e.g. the greater the affective value or utility of one's act for another, the greater the normative value of the act (positive or negative). Whereas this system is in all likelihood innately based, proposes Jackendoff, a major component of learning a culture is learning the rules that assign values to particular sorts of action. This, or something like it, is, I propose, the system that could be said to underlie Taylor's human meanings. And this is, I submit, the kind of theory we should strive for when it comes to human values. It is easy to see how loyalty would fit nicely into it.

Taylor claims that in the case of human meanings, the experience does not preceed the expression. However, how would this be cashed out? One of his examples is the attitude of "cool" as adopted by young people. The idea is that it isn't the case that, first, there was a fully shaped attitude, mode of behavior, which was then just christened as being "cool"; rather, the expression "cool" helped the attitude emerge, take shape. But it seems that a more illuminating account of this process is this: there was an *incipient conceptualization* (emotionally coloured) of what it is to be cool, which additionally crystallized when the term "cool" was coined (rather, transferred metaphorically from the temperature domain). The notion of an incipient conceptualization seems to be able to bring down to earth, so to speak, the rather mysterious notion of the expression "opening up new domains" and not being preceded by what it designates.

The theme of stressing conceptualization rather than language is the second one I would like to develop here in opposition to Taylor. Taylor talks of the efficacy of discourse in constituting social reality. This is of course reminiscent of Searle's (1995). But that book also seems to talk of language where talk of conceptualization would be more adequate. An institutional fact can be created by language, but also by raising a flag or touching one's shoulders with a sword, or putting a crown on one's head. It is the conceptualization of a physical action counting as the creation of a new institutional reality that is crucial; this conceptualization can be prompted/anchored by a linguistic act, but also by other symbolic acts.

We can continue the theme of conceptualization as being dominant to language in discussing Taylor's account of narrative. He sees the constitutivity of language embodied, amongst other things mentioned above, in the sense-making work of narrative. However, recent work on understanding narrative (cf. Turner 1996, Dancygier 2012) draws on cognitive linguistics (the wokk of Lakoff and Johnson and others), which operates under the premiss that language doesn't encode meaning directly, but is a system of prompts for the construction of meaning (cf. Fauconnier and Turner 2002). As Dancygier elaborates, the story that we reconstruct from a text, the *emergent story*, is the result of a process of *blending* of *mental spaces* (technical terms in cognitive linguistics). Mental spaces are conceptual packages we construct on-line during thinking and speaking. The crucial thing here is that language only prompts us to this, but does not contain explicit instructions on how to do it, so that the resulting construction is much richer than the language used to spawn it. To give an example, dealing with complex constellations having to do with (multiple or shifting) points of view in a story requires building and manipulating a multiplicity of such mental spaces. The moral: language is only the tip of the iceberg, and conceptual structure is what's doing the real work here. And not only is language as a rule only the tip of the iceberg, but it is not even necessary to activate the said cognitive structures—this can be done rather well by means of pantomime or pictures, for instance.

The point of the language vs. conceptual structure discussion could be encapsulated thus: it is not that language "creates new ways of being" for humans; it is rather that human conceptual structure, evolved through natural selection (and possible other mechanisms), and expressible by language, makes us what we are, and has made us such since it appeared some 50 000 years ago.

A few final remarks. Taylor insists that language should be viewed as part of a range of symbolic forms, including dance, music, literature (echoing Cassirer). However, language seems to be rather unique amongst these, perhaps justifying the claim of a categorical difference. For one thing, language is the only system which has both form and content, and has undisputable minimal units which combine both. These are morphemes, e.g. "horse" and "s" in "horses". Dance or literature have no such undisputable units (despite valiant efforts of semioticians to identify them). Second, language seems to have an innate basis which channels its development, one which kicks in almost right after birth, and delivers full-blown language by age three. If this window is missed due to lack of input, the language ability never develops in a normal capacity. Nothing of the sort holds of dance or understanding/producing literature.

Finally, Taylor claims that language develops in the context of "communion", i. e. emotion-infused joint attention. This is undoubtedly true, but it is very questionable whether this proves that language is

essentially shared. Chomsky would say that "language is essentially an instrument of thought. Externalization then would be an ancillary process" (2016: 14). It is as yet unclear, that is, to what extent being exposed to external input really shapes language in the child's mind, or merely prompts it to grow along a genetically predetermined course.

In conclusion, Taylor's book is an insightful, learned, ambitious, and coherent discussion of language, that attempts to offer an alternative to accounts dominating current formal linguistics and philosophy of language. I am just sceptical of the approach it argues for and of the limits it claims for the standard picture.

References

Chomsky, N. 2016. *What Kind of Creatures Are We?*. New York: Columbia University Press.

Dancygier, B. 2012. *The Language of Stories: A Cognitive Approach*. Cambridge: Cambridge University Press.

Fauconnier, G. and Turner, M. 2002. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.

Jackendoff, R. 2007. *Language, Consciousness, Culture: Essays on Mental Structure*. Cambridge, MA: MIT Press.

Miščević, N. and Perhat, J. 2016. *A Word Which Bears a Sword*. Zagreb: Kruzak.

Pinker, S. 1994. *The Language Instinct: The New Science of Language and Mind*. London: Penguin.

Searle, J. 1995. *The Construction of Social Reality*. New York: Free Press.

Taylor, C. 2016. *The Language Animal: The Full Shape of the Human Linguistic Capacity*. Cambridge: Harvard University Press.

Turner, M. 1996. *The Literary Mind*. Oxford: Oxford University Press.

Book Reviews

Peter Lasersohn, *Subjectivity and Perspective in Truth-Theoretic Semantics,* Oxford: Oxford University Press, 2017, 293 pp.

The topics of disagreement and interest in the areas like judgments of taste and beliefs about future contingent have been around on the analytic scene for at least two decades. A dozen years ago Lasersohn has proposed an interesting and pioneering relativist semantics, primarily for judgments of personal taste, but extendable to a much wider domain (see Lasersohn 2005). His views have been amply discussed, by authors like Michael Glanzberg (2007), John MacFarlane (2014), Herman Cappelen and John Hawthorne (2009), Tamina C. Stephenson (2007) and others; he addresses their concerns in the present book. Here, he slightly reformulates his earlier proposal, formulates the new one in much greater detail, extends it to a wider range of phenomena, and places it in the context of linguistic and philosophical discussion at the present moment. Here we shall concentrate on philosophically central issues, leaving, with apologies, all the technical linguistic details aside, except for saying that author's treatment of them looks very impressive. We shall look at several issues: the semantics and pragmatics of predicates of personal taste, at the semantics/pragmatics distinction as seen by the author, and at aesthetic predicates, so that we shall not strictly follow the order of the book itself.

The "Introduction" offers motivation for the whole work, and an important characterization of faultless disagreement. It starts from the standard speech situation between two characters. We regard them as disagreeing with one another, Lasersohn notes. And then comes the crucial point, namely that the absence of error makes disagreement "faultless":

> Yet neither one of them seems to be making an error of fact. We may regard each of them as entitled to his or her own opinion—as fully justified in adopting and asserting that opinion—even though this places them in direct contradiction to one another. (7)

This is valid for a wider class of sentences and their uses: "(...) a sentence expresses a matter of opinion if it is declarative in syntactic form, but gives rise to faultless disagreement when contradicted" (7). This variation in truth is not dependent the possible-world parameter, but on the nature of values represented by value indices, so that the content is not true or false tout court, "but only relative to particular values for these non-world indices" (8). So, beside the monadic truth, we get parametrized notion of

"truth-relative-to-indices", some of which are our value or opinion indices.

In the first chapter, "Subjectivity, disagreement, and content" Laser-sohn specifies as his goal offering "a truth-theoretic semantics for sentences expressing subjective judgment (p. 1). The second chapter, "Dismissing the easy alternatives" is dedicated to popular alternatives such as Indexical and quantificational analyses and expressivism. The third chapter, "Setting the syntactic and semantic stage" presents syntactic assumptions, and discusses classical topics, including pronouns, names, and anaphora. The fourth chapter is dedicated to the grammar of time and space. We reach the central philosophical issues in chapter five, "Basic relativist semantics".

Lasersohn's relativism is quite radical (as has been noted by Dan López de Sa some years ago in his (2011)). Lasersohn has already in his (2005) paper introduced the now context parameter, the "judge parameter"; here is the way he describes it in the present work:

> Contextual parameters other than the judge were assumed to be fixed by matters of fact, of course; so the connection between the context and matters of fact about the practical environment was not entirely severed. Sentences that were purely about matters of fact could be distinguished from sentences about matters of taste in that their contents did not vary in truth value among contexts which differed only in the value of the judge parameter. (92)

Now, Lasersohn wants to "articulate each formal context" into two parts, corresponding to the situation in which an expression is used, and the situation in which a truth value is "judged" (93). And, most importantly, "in sentences about matter of taste, the truth value may vary with both parts" (93). Here are then the innovations:

> The primary changes to be made are:
> (1) denotations will be assigned relative not just to a possible world index, but to a world index and a "perspective" index, where each perspective index itself is identified with an ordered triple of an individual, a time, and a world;
> (2) contents accordingly will be identified not with functions mapping worlds onto denotations, but with functions mapping world-perspective pairs (or equivalently, world–individual–time–world quadruples) onto denotations;
> (3) contexts of assessment will be distinguished from contexts of use, and will supply perspectives to serve as arguments to contents in order to derive denotations, including truth values. (94)

I skip important, but less philosophical material in chapters on attitude predicates and on assertion and pass directly to the characterization of disagreement. Indeed, chapter nine, on "Pragmatics of truth assessment", brings essential material on faultless disagreement, where two people assert or believe contents which contradict each other, "without either one making an error of fact" (209). Lasersohn notes that this kind of disagreement

> does not imply that neither party sees anything wrong at all with the beliefs or assertions of the other. One may regard another person's beliefs or assertions as objectionable—and even *mistaken*—in all sorts of ways which do not involve errors of fact. (209)

I find the formulation puzzling, to say the least: if John sees Mary's assertion as mistaken, how can he prevent himself from seeing that there is something *wrong* with it? Lasersohn probably meant "factually mistaken", but he doesn't say it. Errors of taste are not factual errors:

> What is an error of taste? Crucially, this is dependent on perspective: If I believe that roller coasters are fun, and you believe they are not fun, then from my perspective you are making an error of taste, and from your perspective I am making an error of taste. Objectively, there can be no answer, because the error is of taste and not of fact. (210)

Now, the point of introducing the apparatus was to understand the point of faultless disagreement in judgments of personal taste. How can we characterize the point? It is interesting that the crucial story is placed within pragmatics. The initial characterization Lasersohn offers is cognitive:

> two parties will normally engage in a dispute about a matter of taste only if each of them regards the other as making an error of taste. This in no way represents a retreat from the idea that disagreements over matters of taste are faultless in our original sense, but is simply a clarification of what kind of fault was envisaged. (210)

If we disagree about roller coasters being fun, "then from my perspective you are making an error of taste" and vice versa. Surprisingly, Lasersohn then introduce another explanation, a sociological or socio-psychological one. People debate and quarrel for the sake of practical advantage. If I like roller coasters, I wish that more of them be built, and I praise them hoping we, the roller coaster fans will prevail. The point is practical advantage (211).

This goes ill with the beginning of the account, which is clearly cognitive. So, here I beg to disagree. I find the whole idea of faultless disagreement dubious. Consider the options in relation to a statement of taste, of the form A is Φ. The 1st order options are simple. We can have naive non-dogmatist experiencer who simply claims that A is Φ and that's it. On meta-level, such an experiencer is simply agnostic about further matters: is A Φ for other people, who is right about it, and so on. One alternative, a bit more reflective stance is the dogmatist one: If you don't agree, you just don't know about A being Φ. I think people who do sincerely debate the issues are honest dogmatists, who naively believe they are objectively right. The other option is the tolerant, liberal one: "A is Φ; for me, I mean. How do *you* find it?" On the meta-level, dogmatic disagreement goes well with value-absolutism, entailing that one of the parties is simply wrong, and with relativism. If one is not dogmatist about taste predicates, one should accept that dogmatist is simply wrong; no faultlessness is present. The liberal stance goes well with contextualism. If one is liberal, there is no deep disagreement. So, the idea of faultless disagreement is a myth. In this case, liberalism is wiser than dogmatism.

But note that language is open to all possibilities. The language of taste attitudes is compatible with all three first-order stances: with naive non-dogmatism, with dogmatism and with tolerant liberalism. Particular uses of language can be classified along second-order options, as agnostic, absolutistic, relativistic and contextualist. But the whole business is linguistically correct, syntactically, semantically and pragmatically, so I am doubtful that there is a single correct reading of the use of taste predicates and the like. Our agnostic is linguistically in the clear. The absolutist does not reform language, she is into postulating objective value-properties in the world. The relativist is not making a linguistic mistake, and here Laser-

sohn has to agree. Finally, the contextualist is in clear, as far as language alone is concerned; her description fits the liberal usage perfectly, she may only have problems in theoretical accounting for other options, but not with mischaracterizing language as used by the tolerant liberal.

Let me just mention an issue that raises its head in the same chapter. What is the relation between semantics and pragmatics according to Lasersohn? Pragmatic theory, he claims, explains how contexts of assessment provide particular values for their parameters, and how people go about assessing the truth values of each other's assertions (134). Others would claim that this is done by meta-semantics; If it is so, what job is left to pragmatics? I must say that I am not in clear about the criteria; I am even not certain that there is a clear division in the literature. It would have been helpful if Lasersohn were a bit more explicit about his choices.

Chapter Ten, "Between fact and opinion" is philosophically among the most interesting parts of the book. The central idea is "that certain perspectives may be ranked as objectively better than others" and that there is a theoretical possibility that "certain sentence contents vary in truth value from perspective to perspective, yet also have 'objective' truth values with no relativization" (214). That would point to "a middle ground, between fully subjective matters of opinion, and fully objective matters of fact" (214). I find this particularly interesting because the issue arises in the context of debates about response-dependence, and can be traced back at least to Hume: there is a variation in standards of aesthetic taste, but at the same time we tend to see certain sets of standards as more refined and in fact better, than others. After discussing aesthetic judgment and refinement of taste the author passes to other candidates, for instance claims about future contingent events where later perspectives seem better than the earlier ones, so that there is again a possibility of hierarchy. Next come epistemic modals: here, "some perspectives seem inherently better than others for evaluating the truth of such contents. We may therefore assign such contents 'absolute' truth values in addition to relativized truth values, despite the perspectival variation" (224). He concludes with a fine analysis of seemingly unrelated phenomena, scalar cut-offs too, and derogatory epithets, and finds interesting analogy in the possibility of hierarchies of perspectives.

The last and concluding chapter offers an evolutionary fable about possible sources of perspective assessment, and a formalization of perspective relations in an abstract cognitive space (which reminds one of Gärdenfors and his conceptual spaces).

Let me conclude. The book offers an impressive combination of linguistic and philosophical reflection, enriched by impressive technical logico-linguistic skills. It gives a very wide account of the behavior and meaning of centrally important predicates in natural languages, the ones that somehow point to a reference to the speaker or the judge of the sentence in which they occur. I disagree with the main motivation, namely belief in faultless disagreement, but I find the defense rich and impressive. Lasersohn's systematization of various important predicate kinds is very helpful. I agree that some response-dependent properties allow for objective standards and I hope the moral properties are such; it would be nice if aesthetic properties were. Again, I agree that some predicates and properties don't allow, for

instance, the taste predicates. But the very bringing together of a very wide range of domains from aesthetic, through moral all the way to epistemic modals, and future contingent matters, and offering a way to systematize the phenomena appearing in these domains, is an impressive achievement.

References

Cappelen, H. and Hawthorne, J. 2009. *Relativism and Monadic Truth*. Oxford: Oxford University Press.

Glanzberg, M. 2007. "Context, Content, and Relativism." *Philosophical Studies* 136 (1): 1–29.

Lasersohn, P. 2005. Context dependence, disagreement, and predicates of personal tast", *Linguistics and Philosophy* 28: 643–686.

López de Sa, D. 2011. "The Many Relativisms: Index, Context, and Beyond." In Hales, S. D. (ed.). *A Companion to Relativism*. Oxford: Blackwell.

MacFarlane, J. 2014. *Assessment Sensitivity Relative Truth and its Applications*. Oxford: Oxford University Press.

Gärdenfors, P. 2000. *Conceptual Spaces; The Geometry of Thought*. Cambridge: MIT Press.

Stephenson, T. 2007. "Judge dependence, epistemic modals, and predicates of personal taste." *Linguistics and Philosophy* 30: 487–525.

NENAD MIŠČEVIĆ
University of Maribor, Maribor, Slovenia
Central European University, Budapest, Hungary

Bradley Murray, *The Possibility of Culture: Pleasure and Moral Development in Kant's Aesthetics,* Oxford: Wiley Blackwell, 2015, 160 pp.

To put it simply, everyone interested in Kant's philosophy and/or in the way art and nature connect to our morality and our culture, should read Bradley Murray's book. Reader-friendly and easily accessible even to those who have not spent their lives studying the philosophical giant that is Kant, Murray's book offers an intriguing insight into some of the often-neglected aspects of Kant's aesthetics. Beginning with the simple question, why is it ok for us to pursue aesthetic pleasures provided by art and nature, Murray not only manages to explain the relevance of aesthetic pleasure for our personal development and social wellbeing, but he does so by situating Kant's theory of beauty against a wider background of Kant's works, primarily his anthropology and moral psychology. While most of those who work on Kant's third *Critique* tend to either analyze it in connection to the first or the second *Critique* (i.e. either to Kant epistemology or to Kant's ethics), Murray manages to offer a new look at the third *Critique* by situating it against Kant's accounts of emotions, passions and culture, as developed in his *Lectures on Anthropology*, *Metaphysics of Morals*, *Toward Perpetual Peace* and his other works of more empirical bent. A result is an intelligible, clear, precise and above all informative book which motivates one to take up Kant and see his

aesthetics, as well as his overall philosophical system, in new light. In what follows, I will briefly present a chapter-by-chapter summary of the main claims, and I will end by raising some concerns for Murray's views.

Murray's mission is to show how Kant justifies his claim that pursuing aesthetic pleasure is morally relevant because doing so promotes our capacity to act effectively as moral agents. Therefore, various sorts of anti-aesthetic claims (such as those inspired by Rousseau, according to which marvelling at the beauties of art nurtures in people the pettiness of soul, deters people from engaging with political concerns and thus keeps them in servitude, or is expressive of one's self-indulgent tendencies) should be rejected. Pursuing beauty, devoting our time and resources to beauties and aesthetic pleasures, can help bring about happier life for an individual and more flourishing to the society. As Murray puts it in the Introduction, explaining Kant's view of social order developed in his *Idea for a Universal History with a Cosmopolitan Aim*, people by nature exhibit a kind of 'unsociable sociability' i.e. they want to live in a society but they also want to be free from demands of social cooperation and left to pursue their own desires. Because the two are often incompatible, in that an individual is often faced with a challenge of balancing her desires and the obligations towards society, it is only through moral development that an individual can become socialized and capable of putting aside her inclinations. Equally important is development of one's capacities to exercise reason publically, i.e. to be guided by one's own understanding. As Murray shows, both of these aims—development of personal and social culture—can be served and promoted via the pursuit of aesthetic pleasures.

In the first chapter, Murray explains the centrality of individual's moral development for Kant's overall view of nature as ultimately hospitable for, and at the service of, humanity's moral progression. Passions and inclinations hold people back from acting from duty and keep them committed to pursuing their individual desires. Not only can this be detrimental for sociability generally, but it can eventually lead one to suicide, as it can easily happen that one is no longer able to provide for the material things one has been relying upon for one's happiness. While the pursuit of luxury (that is, pursuit of that which is agreeable to one) can initially be beneficial, in that it keeps one away from pursuing more bodily-based passions, in the long run, it is only the pursuit of aesthetic pleasure inspired by beauty that can indeed be beneficial for moral development. Two main functions of the experiences of beauty are of paramount importance here, claims Murray. First, its ability to cultivate in us the capacity to feel love, as stated in §29 of the third *Critique*, and its capacity to teach us to step back from our inclinations, due to its disinterested nature. The strength, beauty and appeal of Murray's book is in his construction of these two claims from Kant's numerous works, and in his showing their full theoretical and practical implications.

Second chapter is dedicated to the bond between beauty and love. The emotion of love fosters our moral development because it helps us distance ourselves from our inclinations, preparing us to love something apart from any interest. However, the step from enjoying natural beauty to feeling love is only made possible via the emotion of gratitude: as Kant sees it, the expe-

riences of beauty inspire in us a sense of gratitude towards whichever maker, actual or hypothetical, that created such beauty. This, in turn, mobilizes in us a desire to give to others out of love. This chapter is also insightful in explicating Kant's overall view on love and its connection to our morality and social interactions.

Chapter three turns to one of the most contentious aspects of Kant's aesthetic theory, his claim regarding the disinterestedness of judgments of beauty. Connecting Kant's account of disinterestedness as developed in third *Critique* with his account of 'contemplative pleasure' of beauty presented in his *Metaphysics of Morals*, Murray reconstructs the relevance of disinterestedness for moral development. Most significantly, the pleasure of beauty is contrasted to the pleasures of agreeable, which are not only connected to our desires, thus motivating us to pursue them, but foster our tendency to become attached to objects and to want to posses them. Nothing of the kind is the case with pleasures of beauty, given that no one can posses the beauties of nature. Disinterested character of the experience of beauty, understood as a lack of concern or desire for the object or its existence, has often puzzled Kant's readers in that it is not altogether clear how we can pursue the experience of beautiful objects without caring for their continual existence. But as Murray clarifies, central to disinterestedness is that "whatever desire I have for the object to exist is not central to the pleasure I feel" (48). A second aspect of disinterestedness, revelatory of Burke's influence over Kant, is that a desire to understand the object "does not occupy a prominent place in experiences of (...) pleasure" (49). Pleasures of beauty, in other words, are severed from intellectual (or emotional) pleasures that objects of beauty may provide, though this doesn't imply that the pleasure of beauty is radically disinterested. In other words, "it is not a state characterized by the application of absolutely no concepts, and it is not a state in which we attend to a special subjective object" (57). Murray is here careful to situate this aspect of Kant's theory against Kant's division between free and adherent beauty (where a desire to understand can coexist with the pleasure of beauty). Pursuing aesthetic pleasure makes it possible for us to distance ourselves from our desires and inclinations, and to bracket our concerns, and to do so in a pleasurable way which does not make a demand for self denial (as, for example, some other ways of pursuing moral wisdom might, as when the agent is supposed to rationally grasp the demands of pursuing his own duty).

In chapter four, Murray discusses our aesthetic interest in, and pleasure derived from, works of art. Two questions concern him; first, whether pursuing art can be beneficial to our moral development, and second, whether the experiences of aesthetic pleasures triggered by art can be analyzed along the same lines as experiences of pleasure derived from nature (i.e. via the notion of disinterestedness). One influential line of reasoning that Kant engages with concerns responding to Rousseau's anti-aestheticism. Not only can works of art foster problematic moral attitudes, thus distracting us from our moral duties, but it can easily happen that an interest in art is in fact a concern for one's own status and reputation. In other words, art lovers are not after moral development but after indulging their own vanity and status-seeking, concerned only with being envied by others for their good

taste. Such concerns can fuel passions which are detrimental for moral development, particularly a passion for recognition, for self indulgence and for one's status. While Kant is aware of this potentially inhibiting influence of art for our moral interest, he nevertheless thinks there might be good consequences to those engagements with artworks motivated by reputation and vanity: even in these cases, the underlying motivation of a vain person is a desire to communicate, to relate to another human being and to find some common ground, i.e. shared aesthetic taste, with others. Such concerns can contribute to one's moral development, even if one's initial motivation in pursuing art was not morally praiseworthy.

What about disinterestedness with respect to works of art? Because they are primarily artefacts, our engagements with them can inspire a desire to understand them, to know the original intentions of their makers and to know which purpose they aimed to fulfil. Such knowledge however undermines the attitude of disinterestedness, which in turn undermines the possibility of moral development. Nonetheless, Murray suggests, one's experience of artwork can still be disinterested, as long as one is able to undergo a process of abstraction, i.e. attend to the object as it strikes the eye. As Murray argues, this includes taking "a step back in our experience of the object so that this experience is not taken over by our desire to understand the object" (75). The more one is able to do this, the more disinterested one's experience of art is. Disinterestedness is thus a matter of degree. This chapter closes with an analysis of the notion of a genius and its relation to beautiful art. As Kant famously claims, beautiful art is only possible as a product of a genius, that is, as a product by someone who is oblivious to the roots of his art, but endowed with a gift of nature to produce beautiful artworks. However, as Murray objects, this kind of 'metaphysical' reading is problematic, in that one can admire an artwork for its beauty even if the artwork is in fact a product of copying rather than a product of original creation through genius. Murray therefore concludes that Kant's account of creation via genius should be understood 'epistemically'. "Kant's doctrine of genius is to function partly as extension of his account of abstraction" (77) claims Murray, which is to say that one should experience the artwork "as if it were the product of mere nature, rather than as the product of a determinate act of making" (77). Consequently, our experiences of artistic beauty include the representation of the work as a quasi-natural entity. This interpretation is in line with Kant's claim that genius is the innate mental aptitude through which nature gives the rule to art. Murray argues that this interpretation makes it easier for us to understand Kant's claims according to which an artwork appears as nature and doesn't exhibit any sign of having been intentionally created. We seek teleological understanding of natural entities, so our desire to understand the maker's intentions with respect to works of art can be accommodated under our wider teleological estimation. One aspect of artistic creation not easily resolved by Kant's theory is the fact that many of our artworks are not pleasing in light of their beauty (Murray analyses Duchamp's *Fountain* as a telling example) and that many are endowed with ethical dimension.

In the fifth chapter Murray's attention turns to sublime, whose relation to morality is far more intimate than that of beauty. But, wonders Murray,

how exactly does it foster our moral development, and why think that it does, when it is not connected to love but to the sense of respect and self esteem? Another issue with the notion of sublime concerns its mixed nature; it is inherently contra-final and therefore displeasureable, and yet, it can be noted with an expression of approval. To address these issues, Murray first explicates the relevance of respect and esteem for our morality. A respect for the moral law itself is relevant, because humans are never completely free of passions and inclinations which deter us from following the moral law. Respect also matters as a pathological (rather than practical) feeling, designating a feeling of the dignity of human nature, which helps us act with greater impartiality. Finally, respect is also directed towards oneself, i.e. one's personhood. As a form of self-respect and self-esteem, derived from one's realization of, and appreciation of, one's rational nature, this feeling helps us fend off our animal inclinations and other 'worldly' concerns which should seem trivial when recognized as standing opposite to out rationality. Murray then goes on to analyze variations of sublimity, particularly mathematical, to explain the cognitive operation of our minds which are relevant for this experience, and mental states that Kant subsumes under sublimity (such as the feeling of disappointment with humanity). On the whole, Murray concludes that the experience of sublimity can serve our moral development because it is, like the experience of beauty, disinterested and pleasurable. However, Kant's account of sublime is not all together satisfying, as he does not in fact make enough effort to explore the connection between sublimity and art, i.e. artistic sublimity.

The sixth chapter is dedicated to what is perhaps the most challenging issue with respect to Kant's philosophy: that of justifying the pursuit of culture over and above the pursuit of one's inclinations. In other words, why strive toward moral progress (i.e. culture), when not doing so can make for a much more enjoyable life? To answer this challenge, Murray explores Kant's ethical writings, explicating Kant's arguments in favour of pursuing culture. He first explores a set of arguments designed from *Metaphysics of Morals* and the *Groundwork*. In *Metaphysics* Kant claims that pursuit of culture is a duty stemming from that aspect of ourselves that is unique to our humanity, our reason. Because reason induces us to strive toward perfection, we have to strive toward culture. In the *Groundwork*, Kant, developing the three formulations of categorical imperative, discusses the case of a man who contemplates turning his back to his talent. Kant concludes that it is impossible that a rational agent should will that anything like the maxim of enjoyment should become a universal law. Murray is here rather critical of Kant's argumentation, finding it unpersuasive: "Kant does not explain the key moves that he takes to support the relevant claim ... that there is a connection between humanity's being an end in itself and our having a duty to choose to pursue culture over pleasure" (109). Murray then explores Kant's second path to arguing in favour of culture, his argument from the Appendix to the Critique of Teleological Judgment, where Kant claims that a life spent in pursuit of enjoyment at the expense of culture is not compatible with living a worthwhile human life. Kant's argumentation here is complex and in order to present it in its fullness, Murray brings it in connection to Kant's views on the nature and humanity's end, on the connec-

tion between purpose and worth, on reflective and determinative judgement and the regulative use of reason. Murray concludes that this is another unconvincing way to ground Kant's insistence on the supremacy of culture over personal satisfactions, emphasizing two difficulties with the Appendix argument: "those relating to the interpretation of nature, and those relating to the claim that we are to think regulatively of humanity as the ultimate end of nature" (118). However, the Appendix argument is relevant for explaining the "ethical underpinnings of Kant's aesthetics" (118), concludes Murray, in that it provides support to those who are committed to pursuing culture, and it has a valuable role in fending off anti-aesthetics arguments.

Finally, in conclusion, after summarizing the main claims of the book, Murray addresses two further questions: the moral relevance of the experiences of ugliness, and the status of the empirical elements within Kant's aesthetics. With respect to the possible moral relevance of ugliness, Murray only sketchily hints at the possibility that ugliness, which triggers displeasure, might in fact be connected to our feelings of hatred and ingratitude. These, in turn are likely to make it harder for us to resist our inclinations and consequently, to make us self-cantered and therefore isolated from others. Regarding the empiricism in Kant's aesthetics, Murray emphasizes the role that Kant's pragmatic anthropology (i.e. his view of culture, which rests on his views concerning what human beings tend to be like, that is, what they make of themselves) plays in his elaboration of the connection between aesthetics and morality. Murray is here primarily concerned with justifying the inclusion of empirical claims into a philosophical investigation, particularly into those such as Kant's—recall that Kant's aesthetic theory is primarily concerned with an a priori account of the justification of aesthetic judgments. However, concludes Murray, "although Kant is against empiricist accounts of *aesthetic judgments* such as Burke's, this does not mean that Kant's *aesthetic theory* rests on theses that are wholly a priori" (134, italics original).

Given the complexity of Kant's overall philosophy, combining his anthropological writings with his aesthetics in order to make the bond between aesthetic pleasure and moral development stronger is certainly not a small task. Murray is to be complimented for his skills in systematising, as well as for the detailed and meticulous analyses he conducts with respect to each of Kant's claims he scrutinizes. It is worth pointing out that Murray's analysis of the relation between aesthetics, anthropology and ethics in Kant's philosophy is not severed from other Kantians' positions, and while he does not engage with polemics and debates in the main text, the footnotes to each chapter provide for insightful pointers on views of other scholars and Murray's position compared to theirs.

There is however always a problem of choice, and one can wonder why Murray neglects taking into account several issues that figure prominently in Kant's account of the connection between aesthetic judgment and culture. For example, surprisingly little is said regarding the sociability, and the interconnection of *sensus communis*, sociability and communication (particularly regarding Kant's treatment of these in §§ 39–40 of the third *Critique*). Another aspect of Kant's third *Critique* suspiciously absent from Murray's analysis is a distinction between (and the implied relevance of both) empirical and intellectual interest in beauty. Perhaps less important for the ques-

tion of sociability, but significant for how one comes to appreciate beauty, is Kant's discussion of the ideal of beauty (§ 17),which is also surprisingly neglected in Murray's reading. The book could also profit if more space was given to a topic that Murray only sketches, namely, the connection between ugliness and hatred. While it would be interesting to hear more on why someone would deliberately go after experiences of ugliness (particularly if this was considered as a variant of Hume-inspired questions regarding the unpleasant feelings that are part of our experience of viewing tragedy), it would be interesting to see if the experiences of ugly (i.e. pursuit of aesthetic displeasure) might somehow figure in the explanation of the aesthetics of shock, or in pursuits of 'ugly' or 'painful' art. The narrowness of Murray's approach of course parallels the narrowness of Kant's theory of art, as grounded in the 18th century conception of art, but it is worth pointing out that Murray occasionally makes a welcome effort to evaluate Kant's views from the perspective of some contemporary artistic trends and practices.

Irrespectively of these 'omissions' (which do not necessarily hamper the overall insightfulness of the book), I do have some worries regarding Murray's analysis. First, I wonder how plausible his account of disinterestedness (with respect to nature as well as to art) is. Of course, Kant himself, as often emphasized by his commentators, shoots for the moon and famously misses, when he demands that aesthetic judgment be disinterested. Murray's solution, according to which what suffices for disinterestedness is that a desire for the object's existence not be central to the pleasure one feels, while theoretically satisfying, might be tricky from the practical perspective: is it really possible for one to gain such a clear perspective on one's pleasure to say with certainty which aspect is central, and which only secondary, to the experience? With respect to art, Murray's claim that disinterestedness is a matter of degree dependent on how far one is capable of abstracting, seems more convincing with respect to some art forms than with others. While one may marvel at the form of the statue or a symphony, it is hard to understand how we might approach literary works or works of narrative art in such a way. The tension here is as much a problem for Kant as it is for Murray.

Second, Murray's analysis of aesthetic ideas seems superficial. Murray makes a valid point in raising concerns regarding the metaphysical account of aesthetic ideas—we can indeed marvel at the beauty of a well crafted copy and mistake it for the original. However, while Kant can be criticized for having missed this possibility, it is worth remembering that his account of aesthetic ideas is primarily put forward to explain creation of art, not its reception—he barely says anything explicit regarding the audience's take on aesthetic ideas. Kant however makes a substantial effort to explain the relevance of aesthetic ideas for a moral development, and Murray is completely silent with respect to it (although explaining the moral relevance of our aesthetic pursuits is his prime concern in this book). Kant's claims in §52, where he urges beautiful art to be connected with moral ideas provides, on my understanding of Kant's aesthetic ideas, for a firm connection between art and our moral development. It is disappointing that Murray ignores it altogether.

IRIS VIDMAR
University of Rijeka, Rijeka, Croatia

Ronald J. Herring, *The Oxford Handbook of Food, Politics, and Society,* Oxford: Oxford University Press, 2015, 904 pp.

The Oxford Handbook of Food, Politics, and Society, edited by Ronald J. Herring, is the selection of texts in which authors approach the given topic from various perspectives. The handbook consists of five parts: Production: technology, knowledge and politics (I); Normative knowledge: ethics, right, and distributive justice (II); Nature: food, agriculture, and the environment (III); Food values: ideas, interests, and culture (IV); Global meets local: contestation, movements, and expertise (V). In this rich variety of texts that give justice to the complexity of topic, I tried to choose the texts directly addressing the food/epistemology or ethics intersection.

On the topic of food, as the editor himself says, there are three main questions—*what* is to be produced; *how* is to be produced, and how is to be *distributed* (7)? Answers to these complex questions are going to be even harder to give if we take into consideration the fact that food is becoming the burning issue of our time because as a humanity, we are facing the problems of global warming, climate change, population growth, pollution of the environment, and the specific demands on the question of food quality and information, coming from the consumers itself. Let me pass directly to issues of ethics.

Michiel Korthals in his text "Ethics of Food Production and Consumption" nicely shows the role of ethics in contemporary system of food production and consummation. He sets out seven domains in which food and ethics intersect. First domain focuses on the problem of global hunger and malnutrition and Korthals points out that questions concerning those issues are ethical questions (Korthals 2015: 234). Furthermore, who is responsible for the malnutrition (lack of necessary micronutrients) as a result of the current food system; problem present also in the wealthier parts of the world, asks Korthals (ibid.). Thirdly, he raises the question of consumption of animals as food. Namely, hand in hand with animal husbandry, questions of sustainability, soil and water pollution coming from the use of chemicals, deforestation and abuse of antibiotics in animal factories, occur as ethical questions (Korthals 2015: 235). An old ethical question is still present—in intensive, industrial production, voiceless animals experience pain and suffering in the terrible treatment in which they are treated as mere objects (Korthals 2015: 235). In addition to the questions of animals, there are huge questions concerning agro-corporations that globally standardize the production, decrease biodiversity and risk the outbreaks of diseases and pests (Korthals 2015: 235). In this domain there is also an issue of food commodification, food price and distribution—these economic issues have ethical implications (ibid.). Maybe the most controversial question is the question of biotechnology and genetic modification which brings the issue of distrust of the public towards the governments (Korthals 2015: 236). The last ethical question refers to the consumers who have vast number of ethical dilemmas concerning food and who are perceived, by politicians and scientists, as irrational end emotional and condescendingly and are not seriously taken into consideration (Korthals 2015: 237). On all these questions ethics is trying to answer by its various approaches and concepts.

As I stated above, maybe the most controversial issue within the food/ethic discourse is one concerning biotechnology and genetic engineering. Therefore, there are a number of texts in the handbook concerning those issues. John Harris and Drew Stewart in the text "Science, Politics, and the Framing of Modern Agricultural Technologies" discuss two major developments within the agricultural technology which have opened the Pandora's Box. The first moment was the "Green Revolution" which refers to the development of the "modern" varieties of major cereals, a process seriously started in the 1960s (30). This "modern variations" involve genetic modification but they are not genetically "engineered". The second important moment refers to genetic engineering in agriculture, starting in the 1980s (30). In "genetic engineering", desirable genes are transferred in a laboratory between organisms in order to create desirable traits that are impossible to occur in nature through conventional breeding (47). This recombinant DNA technology produces cultivars called the "transgenics" (47).

In their text, authors are demonstrating the ways these two major developments are articulated and offer us an evaluation of arguments as well as the analysis of emerging controversial issues. "Green Revolution" encompasses the seed selection and selective breeding, mostly of wheat and rice, in order to get higher yields and varieties tolerant of hostile conditions such as drought (46). This kind of "modern varieties (MV)" are cultivated in monocultures[1] and farmers are forced to invest every year money into seeds and more and more money into chemicals due to the fact that chemicals are less and less effective (47). This raises the question of environment destruction, and also, intensive cultivation calls upon excessive use of water (47).

The critique of this kind of agricultural production has united the environmentalists and political Left. Critics say that this kind of food production is based on conquest of nature rather than cooperation and furthermore, the farmers' unique knowledge is extruded by capitalist and centralist control—farmers are becoming dependent upon this technology and they have to buy the seeds and supporting agro-chemicals every year. Due to this centralized and uniformed way of food production, decrease of biodiversity stands as a big issue, as well as poverty and environmental pollution (45).

Furthermore, political Left through the concept of "food sovereignty" claims the "people's right to define their own food and agriculture policy, and to protect and regulate domestic agricultural production and trade, outside the control of big capital and without fear of the dumping of cheap food by third countries" (45). Next big point of critique is "biopiracy"—for thousands of years people are cultivating, breeding and selecting plants and this universal human activity now is in the hands of several corporations which are making profit from it (45). Furthermore, the GMO's are legally patentable which means that they are a subject of intellectual property rights and a profit from them can be protected which means that the whole industry of seed and food is concentrated in the hands of several multi-national companies (53). Looking through the broader socio-political frame, this method

---

[1] Monoculture is a deeply problematic way of food production. Think of the vast fields of wheat plantation. To plant just a single type of plant brings danger of large scale pest and disease expansion, soil destruction, water pollution and the degradation of biodiversity.

of food production Left sees as "technical fix" for much deeper socio-political problems of poor countries such as land and water access which, despite the "green revolution", are not being addressed at all (50). On the other side, some claim that is morally wrong for small numbers of activists to try to deny potential benefits of this method based on genetically cultivated food crops (45).

As authors claim, there is still an open question wheatear real changes resulted from the cultivation of "modern varieties" due to the fact that we are facing an epistemic and methodological obstacle. Namely, evidences from the studies depend on different framings which are more a cultural and political thing then relient on scientific understanding, and transgenetic varieties are subjected mostly to negative framing and consequently. Taken all this into consideration, it is hard to conclude what are the implications of this kind of agricultural model (52). This example brings forth the problem of "epistemic brokers"—individuals who successfully popularize their own reading and framing of a particular issue; they represent a kind of a bridge between scientific facts and the public. But aren't we always dependent on "epistemic brokers" who take the scientific material and frame it, articulate it and translate it to the public, in any field?

The strongest critique of this model of food production is in the risk. Namely, science cannot guaranty that cultivation and/or consummation of GMO want have any negative consequences. "There will always be insufficient scientific evidence to prove the absolute safety of any product (GMO or not)" (54). It seems that uncertainty and open ending characterize these two important moments in conventional agriculture. Uncertainty and open ending stand as most fair "conclusion" due to the fact that there is an individualized moment of framing particular concepts, as well as the problem of risk which whiteness how experimental this model really is, with uncertain and irreversible consequences.

It is impossible not to note the stands of biotechnology experts whose answer to global food problem lies exclusively in biotechnology. One of them is Martina Newell-McGloughlin. In her text "Genetically Improved Crops" she advocates that "green biotechnology" is the key for climate change through greenhouse gas reduction, crop adaptation and protections, as well as yield increase in disadvantageous soils (69). Genetically modified food, according to her, would supply its consumers with desirable macro- and micronutrients which are lacking in different diets across the globe (73). To her, biotechnology offers a new dimension of innovation which is crucial for maintaining and enhancement of food production (89).

Likewise, Alan McHughen in his text "Fighting Mother Nature with Biotechnology" sees biotechnology as the answer to global need for food. He is radical in his idea that there are solely two options—path of technology and "development" or "return to Mother Nature". This "return" stands for extreme idea which McHughen caricatures and where human is portrays purely as biological specie without *ratio* who basically acts like an animal and has one progeny. McHughen neglects whole domain of culture which is one important characteristic of humans—the possibility to create culture. According to him, in this "return to Mother Nature", humans should live in tropical and temperate subtropical parts, without any technology which

implies the return to pre-industrial way of agriculture and life in general (435–436). His claim that "food produced using modern technologies is the safest and most nutritious ever" (444) I find disturbing taken into consideration that modern, conventional industrial agrotechnology is built on "modern varieties", GMO, and chemical industry and is a method which is, thanks to the amount of chemicals and risks, unsustainable and deeply problematic. It is a big mystery why there are just two radical options for McHughen in his text, as well why did he radically caricatured the possible alternative to biotechnology.

As Thomas Larson in his text "The Rise of Organic Foods Movement as a Transnational Phenomenon" shows, besides biotechnology and the "return to Mother Nature" there is also a third possibility—organic approach to food production. "Organic" stands as "a mobilizing frame for a social movement". Namely, "organic" is a global phenomenon, an umbrella term which brings together number of different methods of food production which have in common the alternative approach to agriculture and are based on cultivation without the usage of chemicals. In the "organic" approach the focus is not on the final product itself (e.g. their nutritional value") but on the production itself (741). As Larson states, "organic" in its beginning focused primarily on methods for cultivating the soil but it expanded on the broader aspects such as human's health, sustainability of the production, as well as broader social-justice issues like justice and peace (741).

To conclude with this topic, the approach which swears in biotechnology I found dangerous because it is based on arrogant idea of superiority over nature. Above mention, factor of risk is enormous—science cannot guarantee for possible consequences which can be irreversible. Why risk when there are friendlier, sustainable alternatives? I think that the problem in "biotechnology-pro-GMO method" is in its atomistic, narrow approach. Furthermore, although science is most certain tool which enables us to get to knowledge that is truthful, the lack of scientific "biotechnology-pro-GMO" approach is in the lack of contextualization. Namely, humans are just a small part of a broader picture and every part is dependent on other parts. The problem with the atomistic perspective is that it is limited just on one segment. Human is just one part in the nature and to position itself in hierarchically dominant position, in the same time with unknown potential risks and consequences is arrogant and dangerous. What is necessary is more holistic approach where human is seen as a part of nature. That approach can obtain more sustainable and long-term solutions to problems of contemporary food production. Besides conventional method of food production, there are many alternative ways and methods which are safe and which take better care for the soil, give healthier food and independence for the farmers. At the same time, they do not make profit to particular centers of power and maybe that is the reason for marginalization.

One more interesting aspect of food/ethics intersection, as mentioned above, raises the question of consumers. People ask questions about food and they demand information, as Joseé Johnston and Norah MacKendrick in their text "The Politics of Grocery Shopping" demonstrate. They write about the political aspect of food. Namely, a person has a power to "vote with her/his dollar" and by doing that, he/she directly chooses who to sup-

port (644–645). There is a niche of people who consume food in a "conscious and deliberate" way and with their dollar they choose "to buy 'green', local, fair-trade, and sustainable products in the service of health, social justice, and sustainability" (647).

This leads us to the question of knowledge. How the knowledge is shaped in global context of agriculture, science, and technology, asks Ian Scoones in his text "Agricultural Futures". He analyses the process of knowledge-formation within the panel of experts, farmer representatives, NGO-s, private sector, industry and different institutions. How the body of knowledge is formed, he asks (844)? How to articulate global and local processes; how to include different perspectives? What is the dynamic of power relations? Who gets to be included in the first place and whose voice is heard (844–846)? What is crucial, according Scoones, is to find ways in order to make processes of participation and engagement more "meaningful, democratic, and accountable" (855).

Due to the fact that we, as humanity, are facing number of problems such as growth of population, climate change, environmental pollution and non-sustainable food production, it is time to question all existing settings associated with food production and consummation. Food in its connection with ethics and epistemology is field which is becoming more and more actual due to those problems. "Sustainability" is not just abstract, theoretical concept but it is necessity postulate that has to come as a priority in order to survive as a specie. That is the reason why questions of food within the philosophy are urgent questions of present time and the future.

<div align="right">

ANA SMOKROVIĆ
University of Rijeka, Rijeka, Croatia

</div>

Table of Contents of Vol. XVII

Articles

Book Discussions

Book Reviews

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

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