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 uttereed 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.

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 fruitfully study the psychological underpinnings of compositional semantics. 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 semantics and pragmatics, and there is a tradition within IBS of attempting 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 dispositions 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, reference 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 psychology of rational communicators, then, these sorts of facts about reference 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**

[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 reverse-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) $\llbracket \text{he}_1 \text{ drinks} \rrbracket^{g,c} = \lambda w \, . \, g(1)$ 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]

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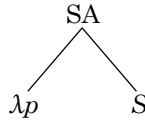
Given this syntactic assumption, what we want is a composition rule that transforms the semantic values of S nodes—i.e., assignment-relativized propositions—into the semantic values of SA nodes—i.e., unrelativized properties of propositions of the kind I described above. This principle should take us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

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

$$\begin{array}{c} \diagup \qquad\qquad \diagdown \\ \lambda p \qquad\qquad\qquad [\![S]\!]^g \\ \qquad\qquad = \lambda w . g(1) \text{ drinks in } w \end{array}$$

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 $[\![a]\!] = \lambda p_{st} . (\exists x^i : \mu(u)(x^i)) \ldots (\exists x^n : \mu(v)(x^n))(p = [\![y]\!]^{g^{x^i/i...x^n/n}})$

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-

---

[32] I lay out this theory in greater detail, and show how to apply it to a wide range of other supposedly context-sensitive expressions, in 'Semantics without Semantic Content' (Harris MSb).

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

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⟧ =
λp_{st} . $(\exists x_e : x$ is called Jones$)(p = \lambda ws$. x 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.