Actual vs. Counterfactual Dispositional Metasemantics: A Reply to Andow
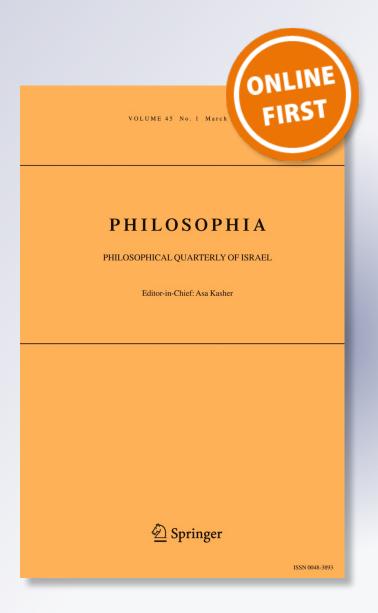
Michael Johnson & Jennifer Nado

VOLUME 45 No. 1 March

ONLINE FIRST

PHILOSOPHIA

PHILOSOPHICAL QUARTERLY OF ISRAEL

Editor-in-Chief: Asa Kasher

🍂 Springer

ISSN 0048-3893

🍂 Springer

Springer

CrossMark

Actual vs. Counterfactual Dispositional Metasemantics: A Reply to Andow

Michael Johnson[1] · Jennifer Nado[2]

Abstract In previous work (Johnson & Nado 2014) we proposed a sketch of a disposition-based metasemantictheory, which has recently been criticized by James Andow (2016). Andow claims, first, that our dispositionalmetasemantics threatens to render the meanings of our words indeterminate, and second, that our viewrisks a 'semantic apocalypse' according to which most of our terms fail to refer. We respond to Andow'scriticism by modifying and expanding our orignial, underspecified view. In particular, we propose that a viewthat appeals to actual dispositions rather than counterfactual dispositions avoids many difficulties that might confront a disposition-based metasemantics - issues even beyond those that Andow raises.

Keywords Metasemantics · Dispositions · Philosophy of language

1 Introduction

A metasemantic theory, as we use the term, is a theory of what makes it the case that an expression means what it does, rather something else, or nothing at all. We have suggested that what plays this role is a speaker's dispositions when that speaker is in possession of 'complete' information. Here we will adopt Andow's name for the view, CIDA (for *C*omplete *I*nformation *D*ispositional *A*ccount). Here is how we characterized CIDA in our initial (2014) paper:

✉ Jennifer Nado
jennifernado@gmail.com

Michael Johnson
Michael.dracula.johnson@gmail.com

[1] Department of Philosophy, Hong Kong University, Pok Fu Lam, Hong Kong

[2] Department of Philosophy, Lingnan University, Tuen Mun, New Territories, Hong Kong

∅ Springer

> CIDA: A linguistic expression E means some object, property, kind, relation etc., X, in the mouth of speaker S, in virtue of the fact that S would be disposed to apply E to X if S had all the relevant information. (Johnson & Nado 2014, 81).

One important notion in CIDA is obviously "relevant information," or "having all the relevant information." We characterized relevant information as follows: "'Relevant information' [regarding an expression E for a speaker S] consists of the facts F that would, were S to be apprised of F, influence S's dispositions to apply E" (p. 81). Possession of 'complete' information with regard to E, then, is a state where there is no relevant information that the speaker lacks—in other words, there is no further information that would alter their application dispositions with regard to E.

There are several ways to motivate the plausibility of this sort of account. In our original paper, we argued that the truth of such an account would imply that application intuitions (roughly, intuitions about, for example, whether 'knowledge' applies to a given case) are generally, though not exceptionally, reliable. This is because people are generally, though not exceptionally, reliable in predicting what their application behavior would be under various circumstances. Given the independent plausibility of the claim that intuitions are generally-but-not-exceptionally reliable, an account that explains this reliability thereby gains strength.

But this is not the only way to motivate a disposition-based metasemantic theory. At bottom, it simply seems as though our dispositions to apply terms must have *something* to do with the facts about what those terms semantically apply to. After all, where do we generally look for evidence about the meanings of terms? To speakers' application behavior with those terms, of course. It seems inconceivable that our application behavior could be *utterly* misleading, such that 'water' in fact applies to all and only, say, toaster ovens. Now, it is obvious that not *all* speaker applications will be correct— sometimes speakers make mistakes. But when they do, it is typically because they are under some sort of misconception; they have mistaken a cow-on-a-dark-night for a horse, for instance. Thus, we're lead naturally to the idea that what matters is not speakers' actual application behavior, but the facts about what they *would* apply their terms to, were such doxastic errors to be remedied. Indeed, it might seem obvious that the semantic application conditions of a term for a speaker must coincide completely with what the speaker would apply the term to, were she under no (relevant) misconception. The step we then take is endorsing the latter as an analysis of the former.

Even further, a disposition-based theory fits quite naturally with the apparent methodology of many of the more famous thought experiments that pepper the metasemantics literature. When we consider Kripke's (1980) Gödel thought experiment, for instance, we consider how we would be disposed to apply the term 'Gödel' were the facts as specified in Kripke's thought experiment, and were we to be informed of said facts. That we would, under such circumstances, persist in applying 'Gödel' to the man we currently apply that name to is taken to be crucial metasemantic evidence. As is the fact that we would not be disposed to apply 'water' to a watery substance when informed that it is composed of XYZ; as is the fact that we continue to apply 'Madagascar' to the island when informed that the name used to be applied to a portion of the African mainland.

Nonetheless, the dispositional metasemantics presented in our previous paper is only a sketch of an account, and crucially, it is underspecified in certain critical respects. This leaves it open to a number of potential objections. Andow (2016) presents two challenges to CIDA—that it may imply that a large proportion of our terms have indeterminate meanings, and that it may imply that a large proportion of our terms are empty. We'll examine these objections a bit later in the paper. There is, we think, a way of interpreting CIDA that leaves it open to objections like Andow's—indeed, we've got our own objections to this interpretation, as well. But we also think that a dispositional metasemantic account in the spirit of our original proposal can be made to succeed, given a few small tweaks. So here is how we will proceed. First, we will explore the various possible interpretations of CIDA, and argue that, given certain plausible interpretations of the definitions we initially offered, CIDA faces problems even beyond those that Andow identifies. We will then propose a new version of CIDA, which we'll call ReCIDA (for 'revised CIDA'), which clarifies the notions left underspecified in our earlier account. Finally, we'll explore Andow's challenges to dispositional metasemantics. We'll pit both the original CIDA and our revised view against Andow's objections, and show that the revised account emerges unscathed even where CIDA fails.

2 Actual vs. Counterfactual Dispositions

We'll begin by looking at a few ways in which our original account is open to multiple interpretations. For convenience, we re-present CIDA and our original definition of relevant information here:

> CIDA: A linguistic expression E means some object, property, kind, relation etc., X, in the mouth of speaker S, in virtue of the fact that S would be disposed to apply E to X if S had all the relevant information. (Johnson & Nado 2014, 81).

> Relevant Information [regarding an expression E for a speaker S] consists of the facts F that would, were S to be apprised of F, influence S's dispositions to apply E.

The first difficulty is with the notion of 'being apprised of' in the definition of relevant information. On one quite reasonable interpretation, one can be apprised of the fact that P and yet not believe P; appraisal on this interpretation is a matter of being told of a fact or presented with it, and does not require accepting it. As we'll argue later, this interpretation leaves CIDA open to one of Andow's objections. But worse, on this interpretation of CIDA, the view is a complete non-starter: speakers' applications of terms to objects easily come apart from the terms' semantic application conditions when speakers are mistaken about the facts. To analyze semantic application in terms of speakers' dispositions to apply terms requires making sure speakers are not mistaken, not merely that they've been presented with the facts. Consider a speaker S mistakenly

applying 'horse' to a cow on a dark night. If helpful stranger H comes along and presents S with the facts, saying "That's actually a cow," S might simply choose not to believe that. How is H, S might reason, any better positioned to see the type of animal that is from this vantage point? However, if S believed that the animal was a cow, instead of being merely presented with the information, S would not apply 'horse' to it.

We admit that we were simply unclear on this point. Thus, we offer a revised notion of relevant information, as follows:

> Relevant Information* [regarding an expression E for a speaker S] consists of all the true beliefs B that would, were S's belief set to be updated with them, influence S's dispositions to apply E.

So consider a speaker who now applies 'cow' to this horse-on-a-dark-night. There is a true belief she lacks, namely that this is a horse, and if this is added to her belief set, she will no longer apply 'cow' to it. Hence this true belief is relevant, by the definition above. We assume additionally a logic of belief revision whereby adding P to S's belief set involves eliminating not-P, and things that entail not-P. Thus S has all relevant true beliefs and no relevant false beliefs. This does not mean she has all true beliefs and no false beliefs, because most true beliefs are irrelevant: whether it's Tuesday or not doesn't affect what one is willing to call 'cow.'

"Relevant information" isn't the only unclear notion in CIDA. There are, in addition, two very different ways to interpret the dispositions CIDA appeals to. The first interpetation grounds meaning in counterfactual dispositions; the second grounds them in actual dispositions. To understand the difference, consider a fragile vase. There are various ways that the vase *is*. Some of those ways are categorical, others are dispositional. The vase *is* two feet tall, and it *is* fragile, that is, it *is* disposed to break, when struck. We might also be concerned, not with how the vase *is*, but how the vase *would be*, if such-and-so circumstances were to obtain. In different counterfactual scenarios, the vase possesses different categorical and dispositional properties than it has in the actual world. The vase is now disposed to break, if struck, but if it were to be strengthened in the appropriate ways, it might cease to have this dispositional property. These are all separate facts about the vase: how it is, how it is disposed to be, how it would be, and how its dispositions would be, in different circumstances.

Now, it's obvious that no one wants a CICA, or a complete information counterfactual account. This would be an account that says that a linguistic expression E applies to an object O in the idiolect of speaker S, in virtue of the fact that S *would apply* E to X if S had all the relevant information—in other words, an account that appeals to mere counterfactuals about a speaker, rather than actual or counterfactual dispositions the speaker possesses or would possess. A few examples will make this apparent. First: most tables are not such that we would call them 'table' be our knowledge ever so great—since others already know they are tables, they don't need us to inform them of this. Second: in the jealous man's idiolect, 'great' applies to many, but he would apply it to none but himself, regardless of what he knew of their greatness. These sorts of cases make

it clear that it is not counterfactual application behavior we are interested in, but dispositions to apply.

This leaves us with the choice between counterfactual dispositions and actual dispositions. The wording of CIDA most naturally suggests counterfactual dispositions: the dispositions one *would* have if one had complete information. As we'll later show, such an interpretation is potentially open to Andow's objections. But we have come to think that an account based on counterfactual dispositions under complete information is problematic for other reasons, as well; thus, we now hold that a dispositional metasemantics must be based on how speakers *are* disposed to apply their terms, *if* they have all the relevant information—in other words, their *actual* dispositions under complete information. It might be a little bit difficult to make out what the distinction here amounts to, so we will start by producing a few counterexamples to CIDA, when CIDA is interpreted as involving counterfactual dispositions. We will then state a revised version of CIDA, and demonstrate how the revised view is supposed to work.

One small matter requires our attention first, however. A canonical disposition is of the form: *to do X if Y* (to produce response X to stimulus Y). So if we want to ask after thing T's dispositions, we'd ask: what is T disposed to do if Y? And we'd receive answers of the form: *to X*, or *to Z*. We could also just ask: what is T disposed to do? Appropriate answers to this question would be of the form: *to X if P*, or *to Z if Q*. However, in talking of speakers' dispositions to apply terms, there is frequently no mention of the relevant *if*-clause. What is S disposed to apply E to? Oh, this, that, and the other. We do, however, think that there is a tacit *if*-clause implied by these claims. We can't speak for others, but reconstructing our own practice, we think the omitted clause looks something like: if S is asked to apply E, and if S then wants to apply E and do so sincerely, and S is sufficiently reflective before applying E, and so on. So when we give counterexamples to the counterfactual disposition version of CIDA, we'll give them to the more explicit formulation here:

> CIDA*: A linguistic expression E means some object, property, kind, relation etc., X, in the mouth of speaker S, in virtue of the fact that if S were to have all the relevant information, then S would be disposed to apply E to X if *S is sincere, reflective, non-reticent, etc.*

Here the italicized portion is taken from Kripke (1979), who needs a device for a similar purpose.[1] Now we turn to why we don't think CIDA*—that is, the fully elaborated counterfactual dispositions metasemantic account—is a good view.

In his (1978), Robert Shope identified what he called the "conditional fallacy." The fallacy consists of using a conditional to analyze some phenomenon, while at the same time ignoring the fact that sometimes the truth value of the conditional itself depends on the truth value of its antecedent. Actual cases are easier to understand than the definition, perhaps. Suppose we say that what it is for X to have a prima facie duty to φ is: if X were to have no other moral obligations besides that to φ, then X ought, all things considered, to φ. Here is Shope's counterexample to this analysis: suppose X has a prima facie duty

---

[1] Kripke discusses a wide variety of cases he means to include in the qualification, and the reader is referred there for discussion.

to teach Y and a prima facie duty to not physically harm Y. According to the analysis, X's first duty amounts to: if X were to have no other duties besides teaching Y, then X ought, all things considered, to teach Y. But, as Shope points out, if X weren't to have the duty not to physically harm Y, then there would have to be some serious reason she didn't, a reason that would likely make it so that X ought not, all things considered, to teach Y. Thus the truth of the antecedent makes the conditional false.

For similar reasons, CIDA* is a bad analysis. If S were to have all the relevant information about some term, S might become a very different person. S might then apply terms differently, not merely because she now has more true beliefs than before, but because of the other changes she has undergone as a result of the process of acquiring that information. Let's think through a case. Suppose that friendship is a matter of attitudes and public comportment. So while Othello might not be Iago's friend, Iago is Othello's friend, because Othello has the appropriate friendly attitudes and comportment toward Iago. Now, Othello speaks English, and uses 'friend' in the normal way (let's suppose); a good metasemantic theory, then, should have it that Othello's term 'friend' applies to Iago. Thus, if CIDA* is to succeed, it must turn out that if Othello had complete information regarding 'friend', he would apply 'friend' to Iago.

But how *would* Othello be disposed to apply 'friend' were he to truly believe all the information that could influence his application of that word, particularly information that could influence his application of the term to Iago? For instance, what would Othello do were he to truly believe that Iago intentionally provided him with false information about Desdemona's infidelity? Surely Othello would *not* then call Iago 'friend.' Perhaps even, he would die of a heart attack on the spot and call no-one 'friend.' Perhaps he would go insane and begin to call all and only Pez dispensers 'friend.' So what Othello *would* do, or *would be disposed* to do, were he to come to truly believe all the relevant information, is not really a good guide to what his terms *now* apply to, while he is on friendly terms with Iago, and of sound mind.

A full-information *counterfactual* dispositions account won't work as a metasemantic theory. But the dispositions aren't what's at fault here. The fault in the account is that not enough is held fixed. We want to consider an idealized version of a speaker S who has all the relevant information, and thus applies terms in ways that avoid *mistakes*. But we don't want the idealization to incorporate the other changes that would result from giving a speaker such information, like changes in her desires or fundamental psychological constitution. The account that's needed, we think, is one in terms of the speakers *actual* dispositions:

ReCIDA: A linguistic expression E means some object, property, kind, relation etc., X, in the mouth of speaker S, in virtue of the fact that S is disposed to speaker-apply E to X, when S has all the relevant information regarding E and is sincere, reflective, non-reticent, etc.

Often, when an object has the disposition to X if Y, then it's also true that were Y to obtain, then the object would do X. But not always—it's well known that the simple counterfactual analysis of dispositions is incorrect, for Shope-style reasons. These involve the so-called finkish dispositions (Martin 1994). A finkish disposition is one
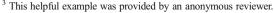
that an object possesses—say, to X if Y—that the object ceases to possess when Y obtains. On our view, Othello's disposition to apply 'friend' to Iago under full information is finkish: it goes away as soon as Othello learns of Iago's misdeeds. But this doesn't mean Othello's expression 'friend' doesn't apply to Iago. It does, because his *actual* disposition is to apply 'friend' to Iago (under full information).

Let us look at the case a little closer. It's given that Othello is now disposed to apply 'friend' to Iago, that is, disposed *if he is sincere, reflective, non-reticent,* etc. But our claim is also that he is disposed to apply 'friend' to Iago, if he is sincere, reflective, non-reticent, etc. *and* if he has all the relevant information regarding 'friend.' Our evidence for this claim is as follows. First, let's be clear that 'friend' is a relational term and Othello is actually applying 'friend' to the pair, <Othello, Iago>. Now note that for every pair x and y such that y is a friend of x, where either x is not Othello or y is not Iago, the following counterfactual conditional is true: Othello would apply 'friend' to the ordered pair <x, y > if he were to truly believe that x has the appropriate attitudes and public comportment toward y (and any other relevant information). We might view Othello as a computational device such that when he detects that an individual x has certain attitudes and public comportment to y, he is willing to apply 'friend' to <x, y>. Only in the case where x = Othello and y = Iago is it not true that Othello would apply 'friend' to <x, y > if he were to truly believe all the relevant information (that is, only in this case do his actual dispositions depart from his counterfactual ones). This is because if he were to truly believe that information, then the grounds for his disposition to apply 'friend' under full information will disappear: that is, his friendly feelings and public comportment. Thus the disposition to apply 'friend' to Iago under full information is finkish: the circumstances that trigger the disposition also cause the disposition to go away.[2]

But since ReCIDA is couched in terms of dispositions, rather than counterfactuals, it still delivers the right result. Since Othello does (actually) have the relevant attitudes and comportment towards Iago, his (actual) application dispositions with 'friend' are such that he is disposed to apply 'friend' to Iago. Again, we might imagine a speaker as a computational device, running a 'program' that determines, given certain information as input, when he will apply a given term. What we want to know is what the 'program' says it outputs when given all relevant facts as input. But, if the speaker *were* to be given all the relevant facts, it is possible that their 'program' *would* change (e.g., Othello would go insane and apply 'friend' only to Pez dispensers)—or it is possible that the relevant input facts *would* be different (e.g., Othello would no longer possess the relevant criteria to trigger an application of 'friend' to <Othello, Iago>). So actual and counterfactual dispositions may often come apart—and it's the speaker's actual dispositions that seem to better track what we take to be the semantic facts.

ReCIDA, then, specifies that relevant information must be actually present in the speaker's 'belief box', and further specifies that the dispositions that matter are actual rather than counterfactual. Before exploring how these changes get us around the worries Andow has raised, we first want to address a potential worry about the model just described. In what sense are the dispositions invoked by ReCIDA truly intrinsic features of the organism, as opposed to, say, my disposition to fly if enchanted by a fairy godmother?[3] The latter disposition isn't really dependent on any features that I

---

[2] It is worth noting—it is not Othello's disposition to apply 'friend' to Iago under *normal* circumstances that is finkish (say, when asked). Only his disposition to apply 'friend' to Iago *under full information* is finkish.
[3] This helpful example was provided by an anonymous reviewer.

possess; a metasemantic account which appealed to such 'extrinsic' dispositions might feel rather unsatisfying or contrived.

Our response here assumes for illustrative purposes a version of classical cognitive science, according to which the brain is a universal computer and the mind is the software that it runs. In particular, minds can be described at a level of analysis that Marr (1982) calls the algorithmic or representational level. This level can be illustrated schematically with flowchart-style diagrams that indicate the order the organism applies its basic operations to effect an output, a methodology sometimes comically referred to as 'boxology.'

A correct boxological analysis of a mind will entail the agent's dispositions to behave in certain ways (or at least, to intend to behave in those ways), given its beliefs and desires. That is, the output intentions will be a logical consequence of the input beliefs and desires and the practical reasoning procedures they are subject to. This algorithmic analysis will also logically entail alternative output intentions when the representations in the belief and desire boxes are modified in various ways. For classical cognitive science, the computational states assigned to an organism are no less 'real' than the states assigned to the organism at a lower level of description, such as a neurological or purely physical description of the organism's brain. Thus, we think it is reasonable to view dispositions to apply when in possession of all relevant information as genuine dispositions of the organism.

We should also here address a related worry, due to Kripke (1982), about the very idea of appealing to dispositions to analyze meanings. Kripke famously argued against a dispositional analysis of rule-following on the grounds that it gives the wrong results, if analyzed counterfactually. What could it mean, he thought, to say that the organism is disposed to produce the sum of two numbers n and m when given them as inputs, if the (representations of the) numbers were too large for the organism to actually store and process? Fodor (1990), somewhat less famously though we think correctly, identified the boxologist's reply: the boxological diagram simply entails the organism's output. Since this level of analysis abstracts away from many of the limitations to be found at the neurological level of analysis, such as the number of cells devoted to memory, or the amount of energy available for information processing, it describes the organism as behaving in ways the organism might well not behave, had it received the input in question. But once again, taking the descriptions at the algorithmic level seriously as descriptions of actual states of the organism, albeit at a higher level than a neurological or physical description, leads us to accept that they are correct descriptions of how the organism does behave when it does have those inputs, even if they are not correct descriptions of how it would behave were it to have those inputs. These dispositions are logical entailments of a correct description of the organism as it is now.

Another way of stating the same Kripkean worry is: 'the totality of my dispositions is finite, being the dispositions of a finite being that exists for a finite time' but 'the idea of meaning something by a word is an idea with an infinitary character' (Boghossian 1989, p. 509). On our conception, however, even finite beings such as us have infinite dispositions, just as the functional characterization of a calculator has infinite entailments regarding its behavior. Boghossian (p. 530) maintains that the burden of proof is on Kripke to show that the algorithmic level description is not an 'illicit' idealization like a tortoise under 'ideal' conditions being able to overtake a hare when it is (ideally) bigger and faster. We think more can be said on this score: the dispositions we're

concerned with involve the organism's current, correct functional description modified only in terms of its inputs (the contents of its belief and desire boxes), not of its program. The tortoise is neither bigger nor faster, but may know better when the starting bell has rung and desire more greatly victory in the end.[4]

Returning to ReCIDA specifically, the first clause in the input specification, namely (a), specifies certain aspects of the speaker's *desires* that are to be altered (if need be) when we calculate the actual dispositions to apply terms that are, we claim, determinative of meaning. For example, the clause specifies that the speaker is sincere, so should the speaker actually have the desire to mislead, ReCIDA suggests that in calculating what her terms apply to, we alter the description of her desire box to remove those desires, and add in desires to the opposite effect. Similar remarks apply to the conditions of reflectivity and non-reticence. Clause (b), on the other hand, specifies aspects of the speakers' *beliefs* that are to be altered. The applications of terms by speakers who are mistaken about the relevant facts are clearly of no relevance to the correct semantic application conditions of those terms for those speakers. A speaker applying 'cow' to every member of a field full of horses, simply because she falsely believes them to be cows, is not engaged in behavior that counts as evidence for what her terms mean. Our analysis takes this observation one step further: we say that not only are the boxological outputs of ideally informed speakers *better evidence* than their outputs when mistaken, but that the outputs of such ideally informed speakers (additionally idealized in certain desires, as above), are *constitutive of* what those terms apply to.[5]

One important contrast between taking current dispositions to be logical entailments of boxological analyses and taking them to be dispositions that are arrived at by a historical (albeit counterfactual) process of true-belief acquisition, is that on the latter, counterfactual disposition account, secondary causal effects of the true-belief acquisition can lead to finks, as when Othello, learning of Iago's deception, ceases to be friends with Iago. On our conception of actual dispositions, however, we are not concerned with any historical process, actual or counterfactual. The description of the

---

[4] There are further objections to dispositional accounts in Kripke's and Boghossian's work – but these further objections are aimed at the project of providing a naturalistic reduction of intentionality, and concern the difficulty of avoiding intentional language in such a reduction. We are not engaged in that project, nor were we in our initial paper – the project here is merely to reduce linguistic intentionality to mental intentionality. Both CIDA and ReCIDA make use of intentional notions – and intentionally so. (See Johnson & Nado 2014, footnote 13).

[5] An anonymous reviewer worries that our view threatens to become trivial: *of course* the semantic application conditions of S's term T are the conditions under which S would speaker-apply T when S has all the relevant information and isn't insincere, reticent, etc. There are two points to make. First, we're not saying merely that the conditions here coincide, but we're answering the metasemantic question: T semantically applies to O for speaker S *in virtue of* S's disposition to apply T to O under full information. The second point is that if our view is so obvious as to be trivial, this is very surprising, since *every other view* that attempts to answer the metasemantic question is incompatible with even the coincidence claim. For example, on a classical descriptivist account, there is a description that S *now and actually* associates with a term T that determines the extension of T in S's mouth. But this can come apart from what S is disposed to apply T to, when S has complete information, if that information is information S doesn't *now and actually* possess. For example, all of S's current descriptions might place whales in the extension of 'fish' but S might nevertheless be disposed to not apply 'fish' to whales under full information. *Mutatis mutandis* for causal-historical theories, for reference magnetism, and so forth – in each case, that which determines the metasemantic facts can come apart from a speaker's dispositions under complete information. If our view is trivial, then all other metasemantic views are trivially false; and it seems clear to us that they are not.

speaker's belief and desire box are altered as per ReCIDA and the speaker's application behavior is derived. As we'll see later, this will be crucial to avoiding certain of Andow's worries.

3 Path-Dependent Semantic Values

We'll now move to explore Andow's objections to CIDA, and how we think ReCIDA avoids them. Andow's first objection is that CIDA winds up assigning indeterminate meanings to too many terms. Andow argues for this claim using two separate thought experiments; we will treat them separately, as our diagnosis of the two cases is rather different. The first thought experiment invokes an actual historical case drawn from Wilson (1982)[6] to motivate the possibility that a speaker might not have a unique dispositions to apply terms under full information, because presentation of the relevant information in different orders might result in different dispositions.

We'll need to start with a bit of background. Around the year 1820, the term "Grant's zebra" was introduced. Zoologically well-informed speakers of English were disposed to apply this term to (and only to) a subspecies of *E. quagga*, namely, *E. quagga boehmi*.[7] Around the same time, the term "Chapman's zebra" was introduced. Zoologically well-informed speakers of English were disposed to apply *this* term to (and only to) a different subspecies of *E. quagga*, namely, *E. quagga chapmani*. The relevant speakers believed the following falsehood: that Grant's zebra and Chapman's zebra were separate species. They believed the falsehood because they additionally believed two other falsehoods: that these animals did not interbreed, and that they were morphologically distinct. As a matter of fact, around the Zambezi river, the animals do interbreed, and there is a smooth morphological gradation from one subspecies to the next.

Those are the facts. Wilson in addition makes another plausible claim about the case: he says that the fact that "Grant's zebra" now refers to the subspecies *E. quagga boehmi* is a historical accident. In particular, his claim is that if exploration of the region (by English-speaking fauna-seekers) had begun around the Zambezi river, instead of at the opposite extremes of *E. quagga*'s reaches, then "Grant's zebra" would have come to mean *E. quagga*, not *E. quagga boehmi*—presumably, "Chapman's zebra" would not have been introduced at all. Let's suppose that this is true.

Andow presents his critique of CIDA against the backdrop of Jackman (1999)'s "temporal externalism." On Jackman's view, the present meaning of our terms can be indeterminate. In particular, Jackman assumes that for some terms, we defer to *future* experts. If the future is open, then it may be the case that in alternative future timelines, the experts say different things. It follows for Jackman that at the present moment, the present meanings of our terms are indeterminate. This is the analysis Jackman gives of

---

[6] Jackman (1999) adapts this case from Wilson (1982).
[7] Wilson, Jackman, and Andow all use "*E. burchellii*" rather than "*E. quagga*," but we've updated to the current nomenclature. This is actually rather appropriate, because the fact that "*E. quagga*" is a name for the species is also a historical accident. It was initially thought that the quagga was a separate species, but recent genetic work has shown otherwise. Thus since the quagga was described first in the literature, that name takes precedence for the entire species. Had Burchell's zebra (the source of "*E. burchellii*") been first described, "*E. burchellii*" would be the name for the species.

Wilson's zebra case. In 1820, there were two possible futures: the one that in fact came to pass, where the experts used "Grant's zebra" for the subspecies, and another one in which explorers investigated the Zambezi earlier, and wound up using "Grant's zebra" for the entire species. Thus, on Jackman's temporal externalist view, in 1820, the meaning of "Grant's zebra" was not settled. (By the way, Wilson also thinks the meaning of "Grant's zebra" is not settled, but for different reasons than Jackman.) Here is then what Andow says about CIDA:

> [CIDA has no grounds for resisting] Jackman's claim that… the discovery of bits of information in *different orders* might affect the meaning of terms as you currently use them… if Jackman is right about cases like the zebra case the relevant individuals have no determinate *dispositions given complete information.* [Andow 2016, pp. 56-57]

First, we'd note that if accepting the indeterminacy of "Grant's zebra" is a problem for CIDA, then it's also a problem for Jackman. But accepting indeterminacy in cases like these is not the end of the world: Wilson lives with path-dependence by adopting partial semantic values, and Jackman deals with it by adopting indeterminate semantic values. And those seem to us reasonable responses to the sorts of cases at issue. That said, if cases like these turn out to be incredibly common, 'biting the bullet' may begin to lose its appeal.

More importantly, we see no reason to believe that this case will result in indeterminate meaning according to CIDA. For this particular case, at least, our original account needs no revision. Let's look more closely at what CIDA says about this case. First, we have the speakers in the past, who introduced "Grant's zebra." Let's call them The Explorers. We have accepted Wilson's (and Jackman's and Andow's) judgment of the case, that had investigation begun around the Zambezi, and thus had The Explorers been in possession of the information that both subspecies interbred and formed a morphological continuum, they (and we now) would use "Grant's zebra" for all of *E. quagga*; thus, it seems plausible that The (actual) Explorers, who did not start their investigation there, nevertheless were disposed to apply "Grant's zebra" to all of *E. quagga*, under full information, at the time of the discovery of *E. quagga boehmi* (henceforth, t_1). So, assuming that is so, according to CIDA, "Grant's zebra" originally applied to *E. quagga*, not just the subspecies *E. quagga boehmi*. Second, we have the speakers of the present, who, let's suppose, possess all the relevant information. They are currently disposed to apply "Grant's zebra" to the subspecies *E. quagga boehmi*. So, according to CIDA, this is a case of meaning change: a name for a species came to name a subspecies, due to the accidents of history. Note that this is not at all unintuitive: The Explorers intended to introduce a term for a species, and the most salient species in the vicinity was *E. quagga*. According to CIDA, it was only later that the entrenchment of usage led to speakers gaining a disposition to defer to precedent—thus resulting in meaning change.

The 'Grant's Zebra' case is indeed a case where information being presented in different orders affects application dispositions – but it is not a case of the right type to threaten CIDA. The discoveries of *E. quagga boehmi* and *E. quagga chapmani* occurred before the discovery of the interbreeding, morphologically similar populations of

E. quagga boehmi and *E. quagga chapmani* near the river; and it is true that if the order of these discoveries had been reversed, The Explorers might have *in fact* applied their terms differently. But this is not what CIDA is concerned with. The Explorers *in fact* applied 'Grant's Zebra' to *E. quagga boehmi* (at t_1 and subsequently); but The Explorers were not in possession of all relevant information. What matters to CIDA is that these very explorers were, at t_1, plausibly disposed to apply 'Grant's Zebra' to *E. quagga generally* when given all relevant information. Suppose a benevolent genie appeared immediately after t_1, and informed The Explorers of the interbreeding population – it's plausible that they would have then applied 'Grant's Zebra' to all *E. quagga*. And it's further plausible that their dispositions would have been exactly the same had the genie instead appeared immediately before t_1, rather than immediately afterwards. The order of presentation of this information does not seem to matter.

This is assuming that the only relevant information is the fact that the animals interbreed and are morphologically continuous. However, we might, taking a cue from Jackman, suspect that relevant information in this case also includes facts about future experts. We emphasize that CIDA, unlike Jackman's view, does not directly appeal to future experts—or indeed any other speakers at all—to determine the content of a term T in a given speaker's mouth. For CIDA, the only things that externally determine content are facts about the case that would affect a speaker's application behavior, were the speaker to be apprised of those facts. Thus, for the case at hand, the future experts' behavior is only relevant to the meaning of "Grant's zebra" in The Explorers' mouths if there was, at time t_1, a fact about those future experts such that awareness of that fact would affect The Explorers' dispositions to apply "Grant's zebra" at t_1.

What such facts might there be? This depends on one's stance on facts about the future. There are two cases to consider. On the first case, we assume the future is not open and there is, at time t_1, a determinate fact about how future speakers will use "Grant's zebra". In such a case, CIDA does not obviously imply any indeterminacy in content – the meaning of "Grant's Zebra" will be determined by whether or not The Explorers are disposed to defer to the usage of the future experts. And it is again unclear that presenting the relevant facts in different orders would result in different dispositions. Imagine, for instance, The Explorers receiving interbreeding information one second after t1 and future expert information two seconds after t1; then imagine the order reversed. It's not obvious to us that this would make any difference whatsoever. On the second case, we assume the future is open, and at time t_1 there is no fact of the matter about what future experts will do. They might apply "Grant's zebra" to *E. quagga* generally, but they also might apply it to *E. quagga boehmi* alone. In this second case, the following seem like all the relevant information: 1) Around the Zambezi river, the animals which are called (in 1820) "Grant's zebra" and "Chapman's zebra" interbreed; 2) There is a smooth morphological gradation between these animals; 3) Future experts might apply "Grant's zebra" to *E. quagga* generally, but they also might apply it to *E. quagga boehmi* alone. Again, the meaning of Grant's Zebra will be determined by The Explorers' dispositions when provided with such information, and again it's not obvious that a change of order would make a difference. CIDA, in its original form, doesn't plausibly imply indeterminate meaning in this case.

4 Intransigence

The zebra case, then, doesn't directly motivate indeterminacy problems for CIDA. Andow's second case, however, potentially does. Andow presents us with the following scenario:

> Take someone who uses the word 'God' and who applies 'God' to the sun (and only the sun) due to some particular religious beliefs. Due to other features of their overall belief set, their faith might be completely insensitive to countervailing evidence—there might be no evidence or argument which could possibly alter their dispositions to apply 'God'. [p. 58]

According to Andow, CIDA is committed to the sun-worshipper having all the relevant information:

> The way we have been understanding 'all relevant information' or 'complete information' is that one is in the relevant state with respect to an expression if there is no new information which would change one's dispositions to apply the expression. Here's another way to say the same thing. To have complete information is to have intransigent dispositions. This is really just a relabeling of the notion of 'complete information' being used. [p. 58]

Now, the objection Andow has in mind is *not* that CIDA gets the wrong semantic value in this case. Rather, the real objection is that there might exist individuals who are 'one step away' from multiple different intransigent positions. For instance, a speaker might acquire an intransigent position when given "the information that leaders in her religion believe that everyone else in the world is deliberately trying to mislead followers of the 'true faith'" (Andow, p. 58); but she also might acquire a different intransigent position when provided with different information.

Consider for instance a near-intransigent Believer listening to Revealer. We might suppose that if Revealer says 'God is the sun,' then Believer will then intransigently apply 'God' to the sun, regardless of any other evidence she subsequently obtains. But we might also suppose that if Revealer instead says 'God is this table,' then Believer will intransigently apply 'God' to the table. And so on, for any other object. CIDA, as Andow interprets it, claims that S's term T applies to O in virtue of the fact that if S were to possess intransigent dispositions with regard to the application of T, S would apply T to O. His point here is that there need not be one such thing: one and the same speaker could have many different things to which she could come to intransigently apply her term T.

We think that the force of this case will depend on the interpretation given to the underspecified term 'apprise'. Recall how we originally characterized relevant information: "'Relevant information' [regarding an expression E for a speaker S] consists of the facts F that would, were S to be apprised of F, influence S's dispositions to apply E" (p. 81). Andow speaks of the sun-worshipper as having particular religious *beliefs*, compelled by features of their *belief set* into a *faith* that is insensitive to *evidence*. In

other words, were they presented with further evidence, they would ignore it. Thus, Andow seems to be interpreting 'apprise' as 'present' – as we've noted, this is reasonable assumption, given our original definitions, though it is not the one we had in mind. But as noted earlier, a dispositional account based on mere 'presentation' with all relevant information is independently problematic. ReCIDA, with its revised notion of 'appraisal', and consequently complete information, is not touched by the sun-worshipper case.ReCIDA defines relevant information as follows: relevant information [regarding an expression E for a speaker S] consists of all the true beliefs B that would, were S's belief set to be updated with them, influence S's dispositions to apply E. On this definition, while it's true that anyone who has complete information regarding T intransigently applies T, it is not true that anyone who intransigently applies T has complete information regarding T. In particular, it's plausible that Andow's intransigent sun-worshipper fails to possess full information.

The sun-worshipper is intransigent because he refuses to accept any additional information he is presented with. People like Andow's intransigent sun-worshipper are easily imaginable, and perhaps even extant. But what's relevant to what their expressions apply to, according to ReCIDA, is not their intransigent dispositions. Rather, it's what their dispositions are, when their false beliefs are replaced with true beliefs. Thus, for the sun-worshipper informed that the leaders of their religion believe that everyone else in the world is deliberately trying to mislead followers of the 'true faith', we must *also* consider what they are disposed to say after adopting the (true) belief that the leaders' belief is incorrect—not to mention numerous other true beliefs about matters theological and astronomical. It is not clear that the sun-worshipper is disposed to continue to apply 'God' to the sun in such a circumstance.

The sun-worshipper under complete information would look something like this: he would truly believe that the sun is just a ball of gas; that the sun did not create the Earth, or the universe, or humankind; that the sun is un-sentient, undeserving of worship, and devoid of benevolence; that it once did not exist, and one day will cease to exist; that it holds no special place in the universe; in short, he would truly believe that it is not, in any way, shape, or form, a deity of any sort. If such a person under complete information persists in calling the sun 'God,' we think there's a good case to be made that their word 'God' just means *sun*.

More generally, for ReCIDA, there aren't multiple ways of becoming intransigent with regard to T as a result of possessing complete information with regard to T. There's only one way things are, and the intransigence we're interested in is the kind people have when they possess true beliefs about the way things are.

The zebra example, we claim, doesn't directly motivate any indeterminacy worries for CIDA. The sun-worshipper case does, if we interpret 'apprise' as 'present' – but ReCIDA's reinterpretation of relevant information avoids the problem. However, Andow might at this point accuse us of missing the forest for the trees. Andow's real worry is more general, and goes beyond the examples we've been discussing. Andow is concerned about the general possibility that different ways of receiving complete information might well result in different application dispositions. Consider a speaker S in the actual world. The original CIDA asks us to consider a counterfactual fact about S—if S were in possession of full information, what *would* she apply expression E to? Andow's ultimate worry is that there is no determinate fact about this. Here is another way to see Andow's point. Suppose we adopt—purely for expository purposes—a

possible worlds semantics for counterfactuals, according to which a counterfactual is evaluated by appeal to the 'closest' possible world in which the antecedent is true. Then we might put the worry as follows: perhaps there is no determinate 'closest' possible world where S has full information. Perhaps in many cases there will be 'ties' between a world where S applies E to O and a world where she does not. But if this is a genuine possibility, then it seems CIDA implies that E has no single determinate meaning.

ReCIDA, however, doesn't work this way. ReCIDA does not employ counterfactuals at all, so there is no possibility of 'closest world ties'. Is there another way in which, say, information presented in different orders might lead to different dispositions on ReCIDA? We think there is not. Recall the nature of the dispositions ReCIDA is concerned with—they are consequences of a boxological analysis of the speakers' mind, where her belief box is altered so as to contain all relevant information. This is not a concrete historical process being envisioned: for ReCIDA, 'the order in which the speaker learnt all the relevant information' does not apply. Consider a calculator. We can of course speak in the following counterfactual way—"If I were to press 5, then ^, then 2, the output would be 25". And here it is true that if one pressed the buttons in a different order, the output would change. But we can also speak of the calculator's abstract functional characterization—one of the functions it computes is $f(x,y) = x^\wedge y$. At this level of description, 'the order in which the information was received' doesn't make sense – the function doesn't 'receive' the information at all. To compute the value of the function we don't need temporal, historical facts—we don't need to ask, e.g., whether one first 'filled in' 5 for x and then 2 for y, or first 2 for y and then 5 for x. It's this latter, functional sort of description that ReCIDA appeals to; and thus, the relevant dispositions are ahistorical. Moreover, there is simply no sense to be made of any other conditions under which the information is received—we don't need to consider, for instance, whether the speaker was in a bad mood when she learned the new information. What we look at is simply what the 'function' outputs for an input consisting of all relevant facts.

5 Semantic Apocalypse

Andow's third objection to CIDA focuses not on indeterminacy, but on the possibility of what he terms 'semantic apocalypse'. This is a scenario in which much of our current vocabulary turns out to apply to nothing. As we have gained information about the world, we have often recognized that certain terms are empty. 'Phlogiston' is the classic example, but one might also mention e.g. 'witch', 'ether', 'hysteria' (in its previous medical use), and so on—not to mention empty names like 'Zeus' or 'Osiris'. In a semantic apocalypse scenario, most of our current terms are essentially in the same boat as 'phlogiston'— empty, and doomed to obsolescence. Andow's worry is that CIDA makes semantic apocalypse probable, simply due to the sheer amount of information that is plausibly encompassed by 'complete information'. Being in possession of complete information would result in a radical change in worldview—which, in turn, would presumably render our current vocabulary unsuitable for use within that worldview. The reason why this is particularly problematic for CIDA is that it implies that our terms are empty *now*—in our actual state of ignorance—due to

the fact that the transcendent understanding of the world that complete information would grant us would result in our abandoning said terms.

On one conception, Andow is in fact pointing out that CIDA commits the conditional fallacy. This is supported by passages like:

> It seems pretty likely that there is information out there which would radically restructure the nature of human existence, make us abandon ways of life, abandon technologies, reconsider our values and place in nature, information which would lead us to restructure the political organization of our species, reconsider national boundaries and the 'artificial divisions' which having distinct languages impose upon us. The likely effect of complete information is semantic apocalypse. [p. 60]

First, given the counterfactual interpretation of CIDA, it's likely that Andow is correct. Consider the word "tine." It applies to the pointy things on a fork. Now it's likely (let's say) that there is *some* truth such that, if we knew it, we'd abandon language and go meditate in a cave until we died. One might think that this truth is irrelevant to the semantics of "tine." But recall what information is relevant for a term, according to our initial statement of CIDA: it's any information that changes one's dispositions to apply the term in question. Ceasing to call anything anything ever again is a change in most people's word-applying dispositions. CIDA, as we've argued earlier, is incorrect when interpreted as based on counterfactual dispositions.

On ReCIDA, speakers' current dispositions have replaced their dispositions in counterfactual circumstances. Let's examine whether this change suffices to avoid Andow's worries. Return to the tine case. Once again, we can see the finkish nature of the case. Speakers are disposed to apply "tine" to the pointy things on the ends of forks, if they truly believe all the relevant information, but the ground of this disposition—their current, normal human desires for communication—is pulled out from under them when the stimulus condition is met. If they *were* to truly believe such things, they *would* restructure their desires and go live in a cave. That is, they *would* have different dispositions to the one's they now have. This case, like the Othello case discussed earlier, illustrates why a dispositional metasemantic account needs to be couched in terms of speakers' current dispositions, not their counterfactual dispositions *were* they to truly believe all the relevant information.

However, it's possible that Andow would still claim that semantic apocalypse is a likely consequence of ReCIDA—in other words, he might claim that 'finkish' dispositions aren't the only problem here. After all, ReCIDA would plausibly say that 'phlogiston' and 'ether' were empty in the mouths of (most) nineteenth century speakers; shouldn't we concede that most of our current, twenty-first century terms are probably just as unreflective of the true nature of reality? Perhaps at the end of scientific inquiry, words like 'electron' and 'mass' will be just as obsolete. Here we have two things to say in defense of ReCIDA. First, if 'electron' really is as unreflective of the underlying scientific reality as 'ether', then we are fully prepared to claim that it is (currently, in the mouths of twenty-first century speakers) empty. Second, and more importantly, the vast majority of terms simply aren't sensitive to the results of future science in this way. Suppose we one day discover that twenty-first century physics is as grossly wrong as Aristotelian physics was—and similarly for every other branch of

science. Plausibly, this won't affect our usage of terms like 'father', 'teacup', 'book-mark', and the like—not to mention 'or', 'not', 'if', and so on. So while we accept that many scientific terms might in fact be empty on ReCIDA, we deny that semantic apocalypse is a likely consequence of our account.

Here's how Andow sums up his worry:

[G]iven the likelihood of semantic apocalypse under complete information, CIDA seems to open up a huge undesirable gap between the meaning of our words and our current state. Our meaning no longer seems to retain any connection to our current state in any important sense. [p. 61]

We agree! There's just something wrong about building one's semantic theory on the counterfactual dispositions of our better-informed selves. Those people are perhaps very different from us, with different histories, and different values. The counterfactual dispositions version of CIDA is wrong. Not only does it commit the conditional fallacy, but Andow is right that it divorces CIDA from what he calls the "intuitive motivation" for dispositional accounts, that they ground meaning in facts about how we actually are. ReCIDA, which is concerned with individuals' actual dispositions, we submit, rectifies this shortcoming. It is who we are, and the dispositions we have, now and in this world, that determines what our expressions apply to.

One final issue merits our brief attention. At the outset, we set aside the application of CIDA to explain the evidential status of intuitions, because our primary goal in this paper was to answer Andow's objections to CIDA. But now that we've replaced CIDA with ReCIDA, it's worth seeing if anything has changed on the intuition front. One attractive aspect of CIDA was that it made sense of our use of hypothetical cases - the importance of using hypothetical examples like the Gettier cases is to 'control for' mistakes or ignorance that a speaker might bring to an actual case, by stipulating the relevant information. The intuition a speaker then has to said hypothetical case provides evidence, or so we think, that were this case to obtain, and the subject were to have all the relevant information about it, and be reflective, non-reticent, sincere, etc. she would apply 'knowledge' to it. Nonetheless, it is imperfect evidence – we are, after all, imperfect predictors of our own behavior, including semantic application behavior.[8]

The issues considered in this paper modify the discussion only slightly. In fact, we think that the introduction of actual dispositions rather than counterfactual ones tightens the connection between vignette-responses and application conditions. Othello's response to a vignette regarding Iago is better evidence of his actual disposition to apply 'friend' under full information than it is of his counterfactual disposition, for exactly the finkish reasons discussed earlier. Nonetheless, a gap remains – one's intuitive access to one's actual dispositions is nowhere near perfect. Thus, our view remains moderate on the evidential standing of intuition.

---

[8] See also Johnson & Nado (2016).

References

Andow, J. (2016). Zebras, intransigence, & semantic apocalypse: Problems for dispositional Metasemantics. *Philosophia, 44*(1), 53–62.

Boghossian, P. (1989). The rule-following considerations. *Mind, 98*, 507–549.

Fodor, J. (1990). A theory of content II. In *A Theory of Content and Other Essays*. Cambridge: The MIT Press.

Jackman, H. (1999). We live forwards but understand backwards: Linguistic practices and future behavior. *Pacific Philosophical Quarterly, 80*(2), 157–177.

Johnson, M., & Nado, J. (2014). Moderate intuitionism: A Metasemantic account. In A. Booth & D. Rowbottom (Eds.), *Intuitions*. Oxford: Oxford University Press.

Johnson, M., & Nado, J. (2016). Intuitions and the theory of reference. In J. Nado (Ed.), *Advances in experimental philosophy and Philosophical methodology* (pp. 125–154). London: Bloomsbury.

Kripke, S. (1979). A puzzle about belief. In A. Margalit (Ed.), *Meaning and use* (pp. 239–283). Holland: Dordrecht.

Kripke, S. (1980). *Naming and necessity*. Princeton: Princeton University Press.

Kripke, S. (1982). *Wittgenstein on rules and private language*. Cambridge: Harvard University Press.

Marr, D. (1982). *Vision: A computational approach*. San Francisco: Freeman & Co.

Martin, C. B. (1994). Dispositions and conditionals. *The Philosophical Quarterly, 44*, 1–8.

Shope, R. (1978). The conditional fallacy in contemporary philosophy. *Journal of Philosophy, 75*(8), 397–413.

Wilson, M. (1982). Predicate meets property. *Philosophical Review, 91*(4), 549–589.