On Complete Information Dispositionalism

Mons Nyquist[1]

Abstract

In a trio of recent articles, Johnson and Nado (2014, 2016, *Philosophia, 45*, 717–734, 2017) defend a form of metasemantic dispositionalism, arguing for a novel approach to the "error"-problem, based on speakers' dispositional states under what they call a state of "full information". In this article, I argue that their brand of dispositionalism fails to solve the "error"-problem, because of what I think of as counterexamples to it. In the final sections, I propose a way to amend the theory to shield it from some of the counterexamples, based on the idea that what determines meaning is not only dispositions to apply words under full information, but also dispositions to evaluate one's prior usage, under full information.

Keywords Kripkenstein · Dispositional theory of meaning · Disjunction problem ·
Complete information · Idealization · Meaning change

1 Metasemantic Dispositionalism and the Error-Problem

Metasemantic dispositionalism is a family of views, formulated in response to the Kripkensteinean sceptical paradox, that hold that there is a constitutive relationship[1] between the meanings and reference of our terms, and how we are disposed to apply them to things (objects, properties, relations, kinds etc.) in the world. Traditionally, metasemantic dispositionalism is the view that what a speaker

---

[1]'Constitutive relationship' is deliberately vague. Some theorists take the relation to be that of supervenience, others property identity, grounding etc. Many metasemantic theories are put forward not only as theories of linguistic meaning, but also of mental content. Johnson & Nado restrict their theory to hold only of linguistic meaning.

✉ Mons Nyquist
    monsandreas@gmail.com

1   Department of Philosophy and Religious Studies, Norwegian University of Science and Technology, Sør-Trøndelag, Trondheim, Norway

means by a word is determined by her dispositions to apply the word to things in the world.[2] To a first approximation: For all speakers (S), words (w), objects, properties, kinds, relations etc. (X)

S Means X by w iff S is disposed to apply w to X[3]

The main challenge for dispositionalism has always been that of dealing with counterexamples to this thesis. One of these problems, and the one Johnson & Nado attempt to solve, is the so-called "Error-Problem" (sometimes called the "disjunction problem")[4]: For many, or most, of the words we use, we are disposed to apply them both to things that are in their extensions and things that are outside their extensions. In other words, we sometimes make *mistakes*, and can be disposed to do so, perhaps systematically so. But if words' extensions were fully determined by how we are disposed to apply them, it would be impossible to apply words to things *outside* their extensions; in other words, impossible to make mistakes.[5] In that case, the dispositionalist would be forced to say that all of our applications are automatically correct and is therefore bound to get the extensions of our words wrong, in perhaps most cases.[6] Many, if not all, dispositional theories on the market attempt to solve the Error-problem by providing a set of normal (sometimes called 'optimal', 'ideal') conditions such that dispositions manifested under these are guaranteed to be *correct*, and therefore meaning-constitutive.[7] Consequently, errors are explained as resulting from applications made under sub-optimal conditions.

---

[2] The literature includes, amongst others: Fodor 1987, 1990; Dretske 1981; Millikan 1984; Blackburn 1984; Pettit 1990, 1996

[3] This is rough and general for simplicity. Strictly speaking, when it comes to properties and relations, the thesis is that S is disposed to apply w to their *instances*. Furthermore, words, in this statement, are individuated non-semantically. Any metasemantic theory with reductive ambitions must, in its statement of the theory, individuate linguistic entities like words, expressions, sentences non-semantically, on pain of circularity. Thus, when I speak of such linguistic entities in this article, or mention expressions using single-quotes, they are always individuated by orthography/phonology alone, and not by meaning.

[4] The other problem is the 'finitude'-problem, which is that since speakers are finite creatures, they are not disposed to apply a predicate F to *everything* that is in its extension. How, then, can we say that F still applies to all its instances (Kripke 1982: 26–8)? Johnson & Nado have a response to this as well, but I will not discuss this problem here.

[5] Here, I use 'mistake' to mean simply *application to things outside a word's extension*, and thus to include both "honest" mistakes, and cases of deception, sarcasm, puns, metaphors etc.

[6] To take the standard example in the literature; most Anglophone speakers are not only disposed to apply the word 'horse' to horses, but sometimes also to non-horses (cleverly disguised robots, toy animals, cows on a dark field in the middle of the night). Since dispositionalism holds that our total set of dispositions determine what we mean by our words, it predicts that in this case, we mean something like *horses-or-cows-on-a-dark-night* by 'horse', but this is simply false – this is not what 'horse' means (in English). Intuitively, our application in this case is a mistaken application of a word with its standard meaning, and not a correct application of a term with a wildly disjunctive meaning.

[7] The exception is Warren (2018) who does not require that a solution to the error problem has to specify conditions that *completely* exclude error, only that it does so in most cases.

2 Johnson & Nado's Dispositionalism

In this section, I present Johnson & Nado's dispositionalism, which I will, following them, call "ReCIDA" ("Revised Complete Information Dispositional Account")[8].

2.1 The Basic Formulation

Johnson & Nado's novel proposal is to specify the meaning-constitutive conditions as those in which a speaker possesses "all relevant information":

> 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 and Nado 2014: 81).

Johnson & Nado partly[9] motivate the view by noting that since a main cause[10] of mistaken word applications is lack of knowledge or information about the scene we are describing, allowing us *more information* about what we speak about, will, plausibly, weed out the mistakes, by giving us a chance to see clearly whatever was unclear before. This seems right for many cases of mistakes: If you are initially disposed to apply 'horse' to cows on a dark field, giving you more information about the animals on the field will lead you to retract your application of 'horse' to them, and change it to 'cow'. If you are initially disposed to answer '158758' in response to '559 × 284 =?', giving you the relevant true beliefs about arithmetic will lead you to change your answer to '158756', given that you are competent with 'multiplication'.[11] In addition to weeding out previously falsely held beliefs, allowing speakers "all relevant information" will also give them new beliefs about matters on which they previously were ignorant – that they had no beliefs about.

2.2 All Relevant Information

Johnson & Nado defines "All relevant information" in terms of *potential change* in the speakers' dispositions to apply a given term, such that information counts as relevant iff it would lead to change in our dispositions:

> '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

---

[8] Their original formulation is their (Johnson and Nado 2014). They subsequently modified and updated their view in their (Johnson and Nado 2017) in response to Andow (2016). In this article I take the 2017 updates to their view as the basis of my critique.

[9] The other motivation is that they hold that it can make sense of- and vindicate intuition-based methodology in philosophy. I will not discuss this here.

[10] Importantly, it is not the *only* cause of mistaken applications, as Johnson & Nado are well aware. Lies and deceit, sarcasm and irony, puns, metaphors and so on, are all instances of incorrect word-application; but by contrast to "mere" mistakes, they are *intentional*. Johnson & Nado aim to explain these kinds of error as well.

[11] Speakers who are unable to use a word *w* correctly *even* given «all relevant information», are plausibly, not competent with *w*. However, this presents a problem for Johnson & Nado's view, which I will discuss in section 3.2.

them, influence S's dispositions to apply E. (Johnson and Nado 2017: 720)
[…] Possession of 'complete' information with regard to [a linguistic expression], 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 [a linguistic expression]. (Johnson and Nado 2017: 718).

Relevant information is, then, all and only information that has the effect of altering our dispositions, were we to come to believe that information.[12] In general, if the state a speaker is in at any given point is such as to lead to certain mistaken applications, it is likely that these will not persist given *even further* updates of true beliefs. In the case where no further true beliefs would alter her dispositions to use a word *w*, it is plausible that these dispositions are *correct*, and so, meaning determining.

Importantly, Johnson & Nado intend 'information' in a demanding *factive* sense stipulating that it is only *true* beliefs that count as information.[13] It is, of course, possible that *false* beliefs could alter our dispositions just as well as true beliefs. For instance, coming to believe a false conspiracy theory may cause us to apply some of our terms differently.[14] But including false beliefs with the meaning-constitutive conditions, would, of course, potentially lead to incorrect applications. Thus, in order to exclude incorrect applications, it seems we must require that the information be *true*, as Johnson & Nado do. However, a worry is then that Johnson & Nado, by helping themselves to a notion of *true information*, already do *presuppose* a notion of linguistic meaning or content, given that *truth* is a semantic/intentional notion.[15]

One more preliminary. 'Information' as Johnson & Nado use the word, is not limited to truths about the *world,* in a narrow sense, but includes also truths about our *linguistic practice*; how we (ourselves and others in our community) do use words, and how we

---

[12] Johnson & Nado emphasize that it is not enough that we are simply *apprised* of the information – we can be *apprised* of a given fact F, in the sense of being told it, informed of it, without believing it. In addition, we need to believe the information (Johnson and Nado 2017: 719–720).

[13] Speaking of 'false information' is not obviously a contradictio in adjecto (to me at least), and it seems 'information' is, on some uses, at least, non-factive.

[14] Some people believe that Paul McCartney died in 1966 and was replaced by a look-alike (https://en.wikipedia.org/wiki/Paul_is_dead). These people might be disposed not to refer to the person they believe to be the look-alike as 'Paul McCartney'.

[15] Traditionally in metasemantics, it has been held that the specification of the meaning constitutive conditions cannot itself *presuppose* a notion of meaning or content, on pain of circularity/triviality: The meaning constitutive conditions must be given in non-semantic and non-intentional terms. For instance, we cannot say that the meaning constitutive conditions for 'red' are *the conditions under which we have no false beliefs about the colours of things*. Interestingly, Johnson & Nado explicitly deny that their account is a *reductive* account in the sense of making no use of intentional or semantic notions, saying that their goal is merely to reduce linguistic intentionality to *mental* intentionality (Johnson and Nado 2017: 725 fn4). Thus, circularity/triviality is not a problem for Johnson & Nado, as they do not intend their theory as an attempt to reduce intentionality to something non-intentional. However, one might wonder whether it is possible to accept (parts of-) their theory *even if* one had reductive ambitions. In this case, one might perhaps suggest that requiring that the beliefs or information be *true* is not required, given that cases where false beliefs lead to wrong applications would probably be excluded from the set of meaning-constitutive circumstances *anyway*, as they would be overturned given yet *further* information, and so, would not be present in the final state in which our dispositions change no more.

have used words, and how our linguistic practice has developed, and will develop, over time. This will be relevant for my criticism later on.[16]

2.3 Actual Vs. Counterfactual Dispositions

Johnson & Nado's idealization of "all relevant information" is open to two distinct interpretations. We can view the dispositions under idealization either as *actual* or *counterfactual* (Johnson and Nado 2017: 720). An actual disposition is a disposition that a speaker has here and now, given her present psychology and physical constitution. A counterfactual disposition is a disposition she would have in a possible world where her psychology and physical constitution were markedly different from how it actually is. This distinction is important, according to Johnson & Nado, because of a crucial difficulty they see as arising for the project of idealizing on speaker's disposition: the idealization *in itself*, could, in various ways, lead to changes in speaker's states, such that this would lead speakers to have different properties than they do *here and now*, merely as a "byproduct" of the idealization (Johnson and Nado 2017: 722). The problem is that such idealization could lead speakers to have different dispositions from their actual ones and so, to mean different things by their words than they do here and now:

> 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. (Johnson and Nado 2017: 722).

But this is not what we want; these *counterfactual* dispositions would not be determinative of our current, actual meaning, but instead of some counterfactual meaning. Thus, we need some constraints on the idealization, in order to exclude the possibility that the idealization in itself changes speakers' dispositions:

> 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

---

[16] This is evident from the following: «[…] 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 continue to apply 'Madagascar' to the island when informed that the name used to be applied to a portion of the African mainland* (Johnson and Nado 2017: 718. My emphasis). In the 'Gödel'-case, the truths that influence our dispositions are truths about the *world* (strictly speaking), namely truths about the history of the discovery of the proof, but in the 'Madagascar'-case, by contrast, the truths that influence our dispositions are truths about our past *linguistic practice* – the fact that the name was originally used to pick out the part of the mainland.

fundamental psychological constitution (Johnson and Nado 2017: 722. Italics as in original).

To counteract such cases, Johnson & Nado require that speakers' current states be modified *only* in terms of their having all relevant information. Should their desires or fundamental psychological constitution change as a byproduct of the idealization, these are to be changed back to what they originally were, pre-idealization.

> […] 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 (Johnson and Nado 2017: 725).

3 Extensional Adequacy

I will now critically assess Johnson & Nado's theory, through counterexamples.

James Andow notes in his reply to Johnson & Nado that a potential problem for Johnson & Nado's view is that:

> [Johnson & Nado's theory] 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. (Andow 2016: 61).

In this section, I spell out some concrete examples of consequences of this «gap», and I argue that they constitute counterexamples to Johnson & Nado's theory. I argue that although the restriction to *actual* dispositions, as defined by Johnson & Nado, is a step in the right direction, in some cases it is still not sufficient to capture our current meanings, as opposed to our counterfactual meanings.

Recall that the job of actual dispositions is to reflect our *current* meaning, since it modifies our dispositions *only* by giving us all relevant information. Johnson & Nado hold that this only has the effect of weeding out our mistakes, and not changing us, and in turn our meanings, in any way beyond this. In this and the next sections, I take issue with this claim. I note that in certain cases, giving a speaker "all relevant information" has the effect of changing her states in such a way that her application dispositions, although they count as actual, by Johnson & Nado's lights, still represent a divergence from speakers' *current* meanings.

3.1 From Ignorance to Competence

We should expect a metasemantic theory not only to explain what makes it the case that words have the meanings they have, but also what makes it the case that words *lack* meanings.[17] On metasemantic accounts that appeal to a distinction between optimal and non-optimal conditions for dispositions to be manifested under, lack of competence is determined by the fact that the speaker under optimal conditions still does not apply a given expression correctly. For example, a speaker who does not understand the word 'addition' is a speaker who under ideal conditions *still* does not answer 'plus'-queries with their sum. Applied specifically to Johnson & Nado's theory, a speaker who does not understand a word *w*, is someone who under full information still does not apply *w* correctly.

But now there is a problem. One effect of giving a speaker "all relevant information" can be the *transition* from lack of a concept to acquisition of a concept - the *onset* of understanding a word. I will now argue ReCIDA has problems accounting for this, as it predicts that a speaker has all the time meant what he does under "all relevant information", and so there is no state of lack of competence, and, so, no point of transition to competence.

Imagine a speaker, Iggy, who is initially disposed to use the word 'orangutan' for various items connected with her breakfast, but in a quite random way. Sometimes she applies it to food items, sometimes to pieces of cutlery, sometimes to people present at the breakfast table. She is disposed to utter sentences like:

- "Today I had orangutans for breakfast"
- "There are lots of orangutans here today"

Since Iggy's usage is so random, we have difficulty saying what she means by the word. It seems her usage is too random to mean anything determinate. One day, upon looking up *orangutan* on Wikipedia and learning that is an extant species of great ape native to Indonesia and Malaysia, and perhaps looking at pictures of them, Iggy's application dispositions for 'orangutan' change. She is from now on disposed to apply it to the species of great ape only.

In this case, the information Iggy found on Wikipedia counts *as relevant information*, since it first, is *true*, and second, it influenced and changed her dispositions to apply the word. We may further suppose that these dispositions will not change further in the future. This implies, according to ReCIDA, that the information Iggy came in possession of counts as "all relevant information" for this word, as she uses it. Thus, her dispositions to use the words after her discovery are meaning constitutive, according to ReCIDA, and her dispositions before the discovery were *not* meaning constitutive.

I think what is right to say about the present example is that coming to possess all relevant information has *led* her to competence. At the time of confusion, Iggy was *not* in possession of "all relevant information", and so, according to ReCIDA, her usage of the words at this stage, was not constitutive of what she meant by

---

[17] Johnson & Nado accept that this is a task for metasemantics: "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*" (Johnson and Nado 2017: 717. My emphasis)

them. It was only upon reading the dictionary definitions that she came to be in possession of *all relevant information*, assuming that her dispositions regarding the words change no more in the future.

The problem now is that since that it is only Iggy's *post*-discovery dispositions that are meaning constitutive, ReCIDA predicts that she has all along meant what she does *now*, post-discovery – she has all along meant *a species of great ape* by 'orangutan'. But intuitively, before the discovery she was *not* competent with the word – she was *incompetent*, because her usage was terribly confused. The problem is that ReCIDA predicts that she *was* competent all along, even at her confused stage, since it is her post-discovery dispositions – those she has under "all relevant information" – that determine what she means, and has meant all along.

Now, the crucial question is, are Iggy's post-discovery dispositions *actual* or *counterfactual*? If it could be argued that they are counterfactual, ReCIDA would be out of the woods. It seems that they are indeed *actual*: they are modified *only* in terms of her gaining "all relevant information", and not in terms of any further changes to her desires or fundamental psychology as the result of the apprisal of information: We can suppose that her desires have all the time been to use the words with their standard, public meaning, and not to mislead or to be sarcastic, and so on. Iggy's psychological state is, then, plausibly, the same before her discovery and after, in the relevant respect. Thus, it is the post-discovery dispositions that are meaning-determining for her use of the words.

The problem for ReCIDA is to make sense of first, the lack of competence and, second, of the transition from lack of competence to full competence. It is forced to describe the state in which Iggy applied the word more or less randomly as one in which she meant what she *now* means by the word: Even at the initial stage where Iggy applied 'orangutan' randomly, it meant a species of great ape, but this is just false. In this case, ReCIDA fails to capture the fact that Iggy meant *nothing at all* by the word.[18]

One could perhaps argue, on a social externalist basis, that even incompetent speakers, given that they defer to experts or the community at large, are to be ascribed the standard, public concept, despite their incompetence. Thus, the social externalist might argue that even in her "incompetent" stage, Iggy should be ascribed the standard concept, ORANGUTAN, only applying it in all cases *incorrectly*, saying falsehoods. But there are limits to this. The original point of Burge's 'orangutan'-example, on which I am relying (Burge 1979: 90–1), is that ascription of the standard, public concept holds only for cases of *incomplete* understanding, in which a speaker has a modicum of analytically true beliefs involving the concept.[19] It does not hold for a *total*

---

[18] Here, it might be tempting to appeal to Dretske's (1981) notion of a "learning period", and say that it is only in the learning period that a speaker's dispositions are meaning constitutive, not before it. Applied to the present case, one might say that the stage where Iggy applies 'orangutan' randomly counts as being before the learning period, and so, are not meaning constitutive; it is only the dispositions she has in the learning period and thereafter that are meaning constitutive. This would allow us to sidestep the problem of the transition from lack of competence to competence, since the pre-learning period dispositions are simply not being taken into account at all. However, as argued by several philosophers, it is doubtful whether Dretske can non-circularly and non-trivially define what counts as the learning period and what does not.

[19] Burge: "There are also examples of quite radical misunderstandings that sometimes generate reinterpretation. If a generally competent and reasonable speaker thinks that 'orangutan' applies to a fruit drink, we would be reluctant, and it would unquestionably be misleading, to take his words as revealing that he thinks he has been drinking orangutans for breakfast for the last few weeks." (Burge 1979: 90–91)

lack of understanding, as when a speaker has very few or even none analytically true beliefs involving the concept, or possesses few or no dispositions to apply it correctly. Thus, in order for a speaker to be ascribed the standard concept, she has to have some modicum of competence, although it is very difficult to say exactly what, or how much competence this should be.[20]

The phenomenon of attaining linguistic competence with a word on the basis of learning new information from other speakers about the meaning of the word is, I think, ubiquitous. Such competence with a word on the basis of learning new information, is often achieved quickly and effortlessly, and on a quite meager basis of new information. For instance, learning words of a foreign language often involves learning that a word in the new language means the same as a word from one's native language. Clearly, this counts as information, as defined by Johnson & Nado. But then, the problem is that since the information leads to semantic competence it gives the wrong verdicts about the state of incompetence.

I should be clear that the semantic competence I have in mind in this example is *semantic* meaning, and not speaker meaning.[21] One interpretation of my case is that though the *speaker* intends to mean, for instance, breakfast items, the semantic meanings of these terms are nevertheless the communal ones. However, I hold that for the speaker to be ascribed the semantic meaning of these terms, some modicum of competence is needed. As I do not think the speaker in my example has such competence, they should not be described the standard meaning, as the semantic meaning. Neither should they be ascribed a speaker meaning.

3.2 Meaning Change

There may be changes in a speaker's dispositions to apply an expression E that lead to a *change* in E's meaning, but this is something Johnson & Nado's dispositionalism has trouble accounting for, since it *identifies* a word's meaning with speakers' dispositions to use it under the conditions where there is no information apprisal of which would lead to a further change in the dispositions.

---

[20] An anonymous reviewer points out that one might take issue with this, and insist that as long as the speaker has deferential intentions towards other speakers, this is sufficient for her to be interpreted as using the public language word, no matter how wrong or ignorant she is about the word's meaning. Here, it should be noted that this is not generally taken to be the case. Part of Evans' story about 'Madagascar' is that Marco Polo had deferential intentions towards the usage of natives of the African country, but nevertheless it seems wrong to say that he referred to the part of the mainland with the name. If we hold all that is required to mean what the community does is deferential intentions, the Madagascar case will not be a counterexample to the causal theory of reference. Further, as argued by Evans (1973) and Dickie (2011), participation in a causal-historical chain of reference is blocked if one does not associate the correct *sortal property* with a name, even though deferential intentions are in place. Evans argues that if one takes 'Anir' to refer to King Arthur's burial place rather than his son, one will not succeed in referring to the son. If they are right about this, one could perhaps argue that Iggy does not succeed in meaning orangutans by the word as she does not associate the right sortal property with the word. Finally, Burge himself argues that minimal competence requires ability to use a term "sufficiently well": "[…] the notion [of cognitive value] marks the minimum competence necessary for ratiocination specifiable with the term. To be attributed the "concept" of sofa - and to have one's attitudes literally and correctly specified with 'sofa' in oblique occurrence - one must be able to use the term 'sofa' sufficiently well to be attributed mistakes and true beliefs about what sofas are." (Burge 1986: 717). The right thing to say seems to be that even though Iggy intends to defer to other speakers, this deference is not successful as she lacks the requisite competence.

[21] Thanks to an anonymous reviewer for asking for clarification on this point.

It may well be that our dispositions, somewhere along the way to their final state, in which they will change no more, intuitively do lead to a *change* in a word's meaning. But, if this is the case, it cannot be right to equate E's meaning with that state of information in which there are no further changes in S′ dispositions to apply E, since we now are unable to say that E's meaning has changed, but must say that E has all along meant what is determined by the dispositional final state. Clearly, in order for a word to change in meaning, it has to be the case that it has one meaning at one point in time, and another meaning at a different point in time, yet CIDA is unable to capture this, since it identifies a word's meaning with its *final* disposition applications. Take an example:

The use of the word 'planet' (or at least astronomical experts' use) changed in 2006 when IAU adopted a definition of the word such that:

> A planet is a celestial body that (a) is in orbit around the Sun, (b) has sufficient mass for its self-gravity to overcome rigid body forces so that it assumes a hydrostatic equilibrium (nearly round) shape, and (c) has cleared the neighbourhood around its orbit. (IAU. (n.d.). Definition of 'planet' https://www.iau.org/public/themes/pluto/).

Imagine a speaker, Ziggy, who is disposed initially to use the word 'planet' to apply to several heavenly objects visible in the sky, one of them being Pluto, to which he has applied 'planet' a good number of times during his life. Assume that Ziggy is generally interested in keeping his use in tune with that of the scientific experts in his community. Suppose that, upon learning of the new IAU definition in 2006, Ziggy is disposed to align his use of 'planet' to that of the definition. He thinks of the new definition as being scientifically more useful, more practical, and that scientists should have the last word in cases like this. In this case, Ziggy has

- one set of dispositions to apply 'planet' before 2006, which includes application to Pluto
- a distinct set of dispositions to apply 'planet' after 2006, which excludes application to Pluto

According to ReCIDA, Ziggy's learning of the IAU definition counts as learning "relevant information", since it is information that influences his dispositions. Next, we assume that his dispositions for 'planet' will not change further after 2006. As such, his dispositions have reached their "final state", which, according to Johnson & Nado, is meaning constitutive. Moreover, the change in Ziggy's 'planet'- dispositions came about as a result of learning *new information*, namely, learning of the new IAU definition. Given these assumptions, ReCIDA predicts what S means *now*, and *has always meant* by 'planet', is determined by his post-2006 dispositions.

All of this implies that according to ReCIDA, Pluto, for Ziggy, has *never* been in the extension of 'planet', since he is not disposed to apply 'planet' to Pluto under "all relevant information". In other words, ReCIDA predicts that his earlier applications of 'planet' to Pluto were *uniformly false*, and so, not meaning determining. But if this is the case, ReCIDA is forced to say that in this case, there is no meaning-change, but simply that 'planet', for Ziggy, has all the time had the extension on which it excludes Pluto. The problem is that, intuitively, this is a case of meaning change. The upshot is

that in order that say that the word 'planet' did correctly apply to Pluto in the past, one needs a different account of meaning than the one that Johnson & Nado gives.[22]

If it could be argued that Ziggy's dispositions to apply 'planet' after learning of the IAU definition are *counterfactual*, ReCIDA would be safe. But also in this case, it seems that Ziggy' dispositions are modified only in terms of his belief system being updated with "all relevant information". Plausibly, it is not the case that his psychology or desires are changed as a *byproduct* of the update. His psychology and desires, plausibly, have remained the same. Thus, his dispositions to apply 'planet' are, by definition, actual, and so, meaning constitutive.

One possible response on behalf of Johnson & Nado is that the notion of "all relevant information" is to be relativized to the information that is in principle available to the speaker at the *time of speaking*[23]: What counts as all relevant information for a speaker S with an expression E, is the information that is potentially available to her at the time of speaking that is such as to influence her application dispositions. The speaker's meaning constitutive dispositions are those she will have if her belief system is updated with all the beliefs that are such as to influence her dispositions, *that are available at the time of speaking*. We can imagine a version of ReCIDA that incorporates this modification:

*Re*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 that is in principle available to S at the time of speaking* and is sincere, reflective, non-reticent, etc.

On this basis, Johnson & Nado could argue that Ziggy is in possession of "all relevant information" in his use of 'planet' *both* before- and after 2006. Before 2006, the information about the IAU definition was not even *in principle* available to him, as it is information about the future, and plausibly, the future is not even in principle epistemically accessible to us. Thus, what makes it the case that the meaning of 'planet' changes is that after 2006 different information falls under "all

---

[22] An anonymous reviewer has suggested that we can view ostensible meaning change cases instead as cases of homonymy or polysemy. This would enable Johnson & Nado to circumvent the problem of meaning change that I point to. However, although I fully agree that any metasemantic theory should be able to account for homonymy, it is not obvious that *all* ostensible meaning change cases should really be construed as homonymy. After all, meaning change could just as well be argued to be something metasemantics should account for (there is already a sizeable literature on how semantic externalism can account for meaning change, following Evans 1973 paper). In particular, I do not see the motivation for such an account in the 'planet'-case, for several reasons. The words 'bank' (riverbank) and 'bank' (financial institution) do not have a common causal origin or developmental history, and so, can be seen to be distinct words. However, 'planet$_1$' and 'planet$_2$' do have a common causal origin and history of uses. Furthermore, a difference between this case and 'bank'/'bank' is that in the latter case the two uses exist *synchronically*, while in the 'planet' case, one usage by and large replaced the other. Finally, ordinary language suggest that we do speak of words' changing in meaning. Of course, if a metasemantic account has reductive ambitions, it cannot presuppose a notion of meaning, on pain of circularity, so it cannot define homonymous words as those that have the same orthography/phonology but different meanings. However, there may be other, non-semantic, criteria for word identity available (for instance, causal origin and chain of transmission).

[23] Thanks to Jussi Haukioja for this suggestion. Johnson & Nado's comments on Andow's Zebra-case suggests that they themselves hold this view (Johnson and Nado 2017).

relevant information", than it does *before* 2006. Thus, *both* before and after 2006 is the speaker's dispositions meaning constitutive, and ReCIDA predicts correctly that there is a change in meaning.

This sounds like the right thing to say, for many cases. For instance, many philosophers want to say that natural scientists' changes in usage can represent the discovery of essences of rather than change the meaning of words.[24] For instance, they want to say that it was a scientific discovery that whales are mammals and not fish, and, so, there was not a change in the meaning of 'fish', despite the fact that speakers up until the discovery were uniformly disposed to apply 'fish' to whales.[25] In order to be able to say this, it has to be the case that even before the information was discovered, 'fish' did not apply to whales, even though most speakers up until around the year 1800 applied it to whales. *Re*ReCIDA's notion of "all relevant information" is well-placed to explain cases like this. We can say that the reason 'fish' did not apply to whales, even in the year 1700, is that although speakers at that time were *in fact* not in possession of all relevant information, they were *in principle* in possession of this information; thus, it is plausible that had their belief system, counterfactually, been updated with "all relevant information" (i.e. the relevant information about whales) in, say, 1700, they would have refrained from calling whales 'fish'.

However, it is possible to modify the above example to cause problems for even *Re*ReCIDA. Let us say that although the IAU definition was officially laid down and made public in 2006, Ziggy had been "living under a rock" and did not become aware of it until 2007, when he read about it on an internet site. In the meantime, he applied 'planet' to Pluto a good number of times, saying things like:

- "Pluto is the ninth planet from the sun"
- "Pluto is a planet, not a dog?!!"

When Ziggy became aware of the definition in 2007, he became disposed to no longer call Pluto a 'planet'. Assume, as before, that his 'planet'-dispositions in 2007 and afterwards will not change further in the future. In this case, the information contained in the IAU definition was *in principle* available to Ziggy in 2006. Therefore, according to ReReCIDA, 'planet' as used by Ziggy throughout 2006 did *not* have Pluto in its extension. This means that he used it uniformly *falsely* when applying it to Pluto. Furthermore, when he in 2007 became aware of the IAU definition, his meaning did *not* change, according to ReReCIDA. But the problem is now that this is intuitively a case of meaning change, and that 'planet', as uttered by Ziggy, intuitively did have Pluto in its extension, up until the point when he learned of the IAU definition. I conclude that in order to exclude cases

---

[24] Kripke: "[...] scientific discoveries of species essence do not constitute a 'change of meaning'; the possibility of such discoveries was part of the original enterprise" (1980: 138).
[25] An exception is Laporte (2004) who argues that scientific "discoveries" are in general cases of stipulated linguistic change, rather than discoveries, properly speaking. "Scientists may appear to have discovered that "the guinea pig is not a rodent". But was this conclusion discovered to be true? No, it was not. The conclusion that "the guinea pig is not a rodent" was not forced upon the scientists. Scientists might have continued to insist that "guinea pigs are rodents" in the light of the foregoing empirical and been no more wrong or right about the essence of "guinea pigs" or of "rodents"" (Laporte 2004: 66).

like this, it seems that ReCIDA, in its present form, is not able to adequately characterize lack of competence or meaning change.[26]

4 A Friendly Amendment

In the final three sections, I propose a way of amending Johnson & Nado's theory to shield it from what I see as the counterexamples to it. I suggest that we can hold onto the notion of speakers' dispositions under "all relevant information", but add that what determines meaning is not only how speakers are, presently, disposed to apply terms under these conditions, but also how they are disposed to *evaluate and correct* their previous history of word-usage, under this condition.

The discovery of learning of a term's original usage can make us disposed to issue higher-order evaluative judgments about our own previous usage of the term. What I will argue is that these judgments are, then, meaning and reference determining for our previous uses of the term.[27] I call the account, which is inspired by some remarks from Evans in his paper "The causal theory of names" (1973), 'RDA' ("retraction-dispositional account"). Evans discusses a case in which appraisal of true information about the past can influence our judgments about the referents of names, and will, in some cases, lead us to decide we were just wrong in thinking a given name referred to a person, and so, will influence our application dispositions.

> A youth A leaves a small village in the Scottish highlands to seek his fortune having acquired the nickname Turnip. ... Fifty or so years later a man B comes to the village and lives as a hermit over the hill. The three or four villagers surviving from the time of the youth's departure believe falsely that this is the long-departed villager returned. Consequently, they use the name Turnip among themselves and it gets into wider circulation among the younger villagers who have no idea how it originated. I am assuming that the older villagers, if the facts were pointed out, would say *It isn't Turnip* after all rather than *It appears after all that Turnip did not come from this village*. In that case I should say that they use the name to refer

---

[26] A surprising consequence of ReCIDA is that it gives seemingly easy solutions to semantic identity-puzzles like Frege's and Kripke's belief puzzles.

1. Hesperus is Hesperus
2. Hesperus is Phosphorus

Part of the set-up of the puzzles is that a *fully competent* speaker can be disposed to assent to (1) but not (2). But according to ReCIDA, full competence is identified with the state of all relevant information. But in that state, the speaker is not disposed to withhold the identity predicate from Hesperus and Phosphorus anymore, as she is informed of the identity. Thus, given all relevant information, the puzzles will simply not arise at all. Millians will perhaps see this as a feature of the account as one might suggest that this does adequately explain their intuition that co-extensive names mean the same thing. Fregeans, on the other hand, will count it as a bug, and will say that such a solution does not capture our intuitions that such names figuring in identity statements are different in meaning at the original time of usage.

[27] Of course, this presupposes that «all relevant information» is not restricted to the information that is in principle available at the time of speaking (or at any other time), given that (much) information about the past, plausibly, is not in principle available to speakers.

to A, and in fact, denoting him, say false things about him (even by uttering *Here is Turnip coming to get his coffee again*). (Evans 1973: 23).

The basic idea of RDA is that the discovery of learning of a term's original usage makes us disposed to issue second-order evaluative judgments about our own previous usage of the name. These judgments are, then, meaning and reference determining for our previous uses of the name. Having a self-evaluation dispositions is a matter of being disposed to judge one's earlier usage of a term as either correct or incorrect in the idealized state of all relevant information. This procedure follows a series of steps. First, there is an appraisal of information, in which the speaker is given information about her previous usage history. The speaker will, then, react to this appraisal by issuing a judgment about the correctness or incorrectness of her previous usage. The resultant verdict is, then, what determines the meaning of the word used, by the speaker.

An example of a how such dispositions work, take the following case. It was discovered that the word 'Madagascar' was originally used to refer to a part of the African mainland, and that it began to be used to refer to the island by Marco Polo in the thirteenth century. Upon making discoveries like these, speakers have certain "backward-looking" dispositions to take into account, and evaluate their past actual use of the word. Upon hypothetically making these discoveries, a speaker S has a disposition to either

- see her previous actual uses as correct, despite the fact that this use does not correspond to the original usage – this is what I call a 'retention disposition'
- see her previous actual uses as incorrect – since it does not correspond with the original usage of the name, this is what I call a 'retraction disposition'

Now, the claim of the theory is that these dispositions, to either retract or to retain the actual use of a word upon such a discovery, play a *constitutive* role in determining the reference of the term. It is this: if one is disposed to view the past utterances as true (under full information), they *are* true. If one is disposed to view the utterances as false (under full information), they *are* false. This, in turn, determines whether reference change has taken place:

- if one is disposed to retract and say that one's previous actual utterances containing the word were false, there is no reference change
- If one is not disposed to retract and say that one's previous actual utterance containing the word were false, there is a reference change

Applied to the 'Madagascar'-case, if Marco Polo is disposed, upon discovering that 'Madagascar' was initially used to pick out the part of the mainland, to deem his previous actual usage of the name as *mistaken*, this entails that his usage indeed was mistaken, and, so, there is no change in reference. If, on the other hand, he is disposed to view his usage as correct, it is correct, and there is a change in meaning.

To see how RDA applies to ReCIDA, consider again the 'planet'-case. RDA claims that Ziggy, in the state of "all relevant information" has dispositions not only to apply 'planet', but also to evaluate his own previous usage of the word. I claim this evaluation plays a constitutive role in determining whether a meaning

change has taken place. It is plausible that if Ziggy is disposed to evaluate his own previous actual applications of 'planet' to Pluto (up until 2007) as incorrect, he takes the word *not* to have applied to Pluto *even then*. RDA holds that this disposition plays a constitutive role in determining the extension of the word on Ziggy's past actual usage: given that he takes his previous applications of 'planet' to Pluto to have been incorrect, they are incorrect, according to RDA. Given that his previous applications of 'planet' to Pluto are deemed false, there is no meaning change; 'planet', as used by Ziggy, has never had Pluto in its extension. If, on the other hand, Ziggy is disposed, under "all relevant information", to evaluate his past application of 'planet' to Pluto as instead *correct*, they *are* correct. In this case there is a meaning change, since the word *now* has different correctness conditions, from what Ziggy deemed it to have in the past, given Ziggy's dispositions to refrain from applying it to Pluto under "all relevant information".

The basic idea behind something like RDA is what I take to be the intuitive idea that our *self-corrective* tendencies play a role in determining what we mean by our expressions.[28] ReCIDA captures *part* of this by claiming that it is the dispositions we end up with in the course of our coming to possess "all relevant information". But it does not capture the role of self-correction of our *past* word use. In order to capture this, backward-looking corrective dispositions are also needed. Only then can we make sense of meaning change, as correction or present usage will only give us criteria that specify that the speaker has all along meant what she does under idealized conditions. ReCIDA winds up in problems like the ones I have sketched in the previous sections since it is unable to say that a speaker's words meant different things in the past than they do under the idealization. In order to capture this, backward-looking corrective dispositions are also needed.

That said, I do think that something like RDA should be quite welcome to proponents of ReCIDA, as the theoretical machinery it employs is, in practice, much the same as that of ReCIDA. ReCIDA emphasizes the role of corrective dispositions under idealization as meaning constitutive. What RDA adds to this is simply another *kind* of corrective disposition: the disposition to evaluate one's earlier usage under idealization. Indeed, something like RDA can perhaps be seen to be already implicit in ReCIDA; after all, our ordinary concept of self-correction can plausibly be seen to be a judgment not only about our *present* activities, but also those of our *earlier* selves. Often, when a mistake is pointed out to us, for instance a pronunciation mistake or a factual mistake, say, about the capital of Switzerland, it is quite natural for us to make judgments concerning the incorrectness, not just of our *current* usage ("Ah, I was wrong just now!"), but also of that of one's *past* usage. Upon being apprised of such a mistake, a likely reaction will be something like "Ah, I have been wrong about this my whole life!". Thus, at least the *presence* of retraction dispositions should not be doubted. The basic idea behind RDA is that such judgments are themselves meaning-constitutive.

---

[28] Philip Pettit develops a dispositional theory of rule-following based precisely on the idea that human speakers possess two kinds of disposition regarding words (Pettit 1990, 1996). First, speakers have dispositions to apply words to objects. Second, they also have higher order dispositions to monitor and correct their first-order dispositions. RDA can be seen as a way of applying something like Pettit's theory to the case of reference change.

5 Motivation: Why Not Dispositions at the Time of Speaking?

I have presented RDA and argued that it seems to give the right verdicts in some cases. However, is there something more positive that can be said for it? In particular, how does it compare with competing accounts of the same phenomena? In this section, I pit RDA up against what is the standard dispositional view in the literature.

RDA differs from what I take to be the standard view about meaning change in the dispositional literature, a view that seems also tacitly presupposed by many semantic externalist theories of reference (more on these externalist theories below). This is the following view, succinctly described by Chalmers in his characterization of Carnap's dispositional account of meaning change.

> An expression E undergoes change in meaning between t_1 and t_2 for a speaker iff the speaker's intension for E at t_1 differs from the speaker's intension for E at t_2. If we accept Carnap's dispositional account of intensions, it follows that E undergoes change in meaning between $t1$ and $t2$ iff there is a possible case such that the speaker is disposed to associate different extensions for E when presented with the case at t_1 and t_2 (Chalmers 2012: 205).[29]

Call this the dispositionalist's 'standard view' about meaning change. What a speaker means by w at t_1 is determined by her actual and possible dispositions to apply w at t_1, under idealization (whatever the idealization might be). A change in what a speaker means by an expression occurs when and only when there is a change in the speaker's dispositions to apply the expression, under idealization, from a time t_1 to another time t_2. Johnson & Nado's ReCIDA, as presented above, is not fully fleshed out in this respect, and it can be interpreted either as a version of the standard view, one that takes speakers' dispositions, at the time of speaking, under the idealization of "full information" as meaning constitutive; alternatively, it can be seen as an instance of RDA. I will argue that ReCIDA should be construed along the lines of RDA, as I think it is RDA that is, in the end, able to give the correct predictions on meaning.

The standard view differs significantly from RDA. On RDA, it is not the counterfactual dispositions *at the time of speaking* that determines whether a meaning change has taken place, but rather a speaker's backward-looking dispositions, at a given *future* time, under idealization, to correct her *earlier actual* usage. Given that RDA differs from the standard view, a pertinent question is why we should accept it, over the standard view. Now, I will provide some motivation for accepting RDA and rejecting the standard view. By doing so, I hope to alleviate worries that RDA is ad-hoc,[30] as it might be argued to be given that it seems to be wheeled in purely with the intention of making sense of meaning change. To show this, I will now argue that all forms of semantic externalism need to explain a certain phenomenon: In cases where speakers change their *present*

---

[29] For my purposes in this article, I want to discuss meaning generally and not commit to any particular descriptive semantic account, so we can simply read 'intension' as *meaning*.

[30] Thanks to an anonymous reviewer for raising this objection.

usage upon appraisal of information about meaning from other speakers or experts, there is also the question of the correctness conditions of the speaker's *past* (pre-correction) usage. I will argue that traditional semantic externalism is, in a certain respect, underspecified, and so, cannot, in its standard form, determine the correctness conditions of past usage. There are (at least) two possible accounts of which externalism can avail itself to solve the problem, what I above called the 'standard view' and RDA. I will argue that we should accept RDA as the best explanation, out of the two candidates, of this.

Typically, externalist metasemantic theories appeal to (various forms of) deferential dispositions or intentions as part of their framework. The work done by such dispositions or intentions is to make it the case that the reference of speakers' uses of a word are semantically governed not by the speaker's internal states, but by (features of) the external environment, be it the physical, social or temporal environment. In other words, the role of deference is to "shift the burden"[31] of reference determination from the speaker's internal states to the environment. It is plausible that if no such intentions are present in the speaker, the environment will not play any role in securing reference, for the speaker.[32]

Let us look more closely at how such deferential intentions work in practice. In Burge's 'arthritis'-case (Burge 1979), the patient, Art, initially believes that arthritis can affect not only joints, but also other parts of the body. His doctor informs him that arthritis cannot affect one's thighs. Upon appraisal of this information, Art decides no longer to say that he has arthritis in his thighs, but to use 'arthritis' only to pick out the disease that the medical experts use the word for. As a result, Art's use of 'arthritis' has its reference determined by the experts' use of the word. Thus, what externalism appeals to is a change in usage dispositions upon (actual or counterfactual) appraisal of information (either from other speakers, experts or the environment itself). Which aspects of use do such dispositions have consequences for? Clearly, Art's usage dispositions upon appraisal have implications for the correctness conditions of his *present* use, and also *future* use, of 'arthritis' (unless he decides to change allegiances again).

In addition to this, semantic externalist theories often make a further assumption about words' *past* correctness conditions. In Putnam's Twin Earth-scenario, he claims that already in 1750, years before the true chemistry of water was discovered, speakers used the word 'water' to apply to only H_2O (Putnam 1975). In the 'arthritis'-case, it is very natural to interpret Art's usage disposition upon appraisal to imply also that his *past* uses of 'arthritis' to pick out inflammation in thighs and other parts of the body aside from joints, were uniformly *incorrect*, and that he always had spoken falsehoods in saying e.g. that he had arthritis in his

---

[31] Borrowing some terminology from Cohnitz and Haukioja 2013 who argue that what theory of reference is true of a given term is something that is determined by speakers' dispositions. For a term to have an externalist metasemantics, they claim it necessary for speakers to have such burden-shifting dispositions.

[32] On Kripke's causal-historical picture, this is a matter of intending to use a word with the same reference as that of the speaker from whom one learned the word (Kripke 1980: 96). On Burge's theory, this is a matter of letting oneself stand corrected by either experts in one's community or the community at large (Burge 1979, 1986). A third kind of deference is to the external environment itself (to let reference be up to the environment).

thighs (Burge 1979). In the 'whale'/'fish'-case, it is natural to presume that speakers always had meant 'fish' to exclude whales, and there was no meaning change signaled by the application change. In all cases, semantic externalists assume that speakers' present change of usage signals not meaning change, but a continuation of earlier usage. However, this is not the only possible interpretation of speakers' deference and subsequent change in usage upon appraisal. There are, in fact, two additional possibilities, as I have argued above.

- The past usage can be seen as having different correctness conditions from the present usage (in which case there is meaning change) – Ziggy-case above (3.2)
- This past usage has no correctness conditions (the past usage was meaningless) – Iggy-case above (3.1)

I will now argue that semantic externalism, in its standard form, is in a certain respect underspecified and so, underdetermines which hypothesis is correct. Note that a disposition to change usage upon appraisal can sometimes signal *meaning change* vis à vis past usage, rather than a *correction* of past usage. Above I gave the example of 'planet'. Take the following case, which I think is structurally similar.

> The vikings in Norway celebrated what they called 'Jol' for hundreds of years, a celebration in honour of their gods in Valhalla. When Norway was christianized around the year 1000, the vikings were persuaded by the authorities to celebrate instead the birth of Jesus, but were allowed to celebrate in much the same fashion. The celebration kept the name 'Jol', a name which still remains today ('Jul'). Thus, they began to apply 'Jol' to the Christian Christmas celebration, rather than the celebration of their own gods.

In this case, it seems false to say that the vikings' deference and subsequent change in usage dispositions upon appraisal was one of *correction* of past usage. Rather, what happened was that they *agreed to change the meaning* of the word 'Jol'. If this is right, it seems speakers can agree to change usage *without thereby correcting their earlier usage*: Present semantic deference and conformity (that is, deciding to change usage upon appraisal) could happen for lots of reasons, only *one* of them being correction of prior usage. It could be the desire to communicate efficiently, social or political pressures, the desire not to be perceived as stupid or ignorant, and so on. As a matter of sociological fact, most speakers are probably generally deferential towards the community or experts upon being informed that their usage is non-standard. Since correction of prior usage is just one reason, of many, for presently deciding to change usage upon appraisal of information, such a decision does not necessarily signal that one's earlier usage was incorrect. Thus, a speaker's present deferential decision to change usage *in itself* does not provide enough evidence to adjudicate which of the two alternatives a-b is correct.

(a) There is no change in meaning, and her earlier usage is incorrect.

(b)   There is a change in meaning, and her earlier usage is correct.[33]

Given such underdetermination, simply having a deferential intention, here and now, to mean the same as the community, itself tells us nothing about past correctness conditions. As noted, semantic externalism makes assumptions about past correctness conditions. But the underdetermination means that the most externalism, in its standard form, can explain is the *present correctness* (and future) correctness condition of uses, not past, pre-correction uses. Thus, externalism must supply a metasemantic mechanism that determines the correctness conditions of past, *pre-correction* usage, in order to be able to draw the distinction between changes in usage that signal meaning change from those that signal correction of usage. As I see it, there are two possible metasemantic explanations available to semantic externalism on this point. One is what I labeled the 'standard view' that relies on speakers' corrective disposition *at the time of speaking*, and the other is the backward-looking dispositions of RDA. In what follows, will argue that it is RDA that best explains the present phenomenon, and so, we should accept it.[34]

While I do think that the standard view gives the right results in the Twin Earth, 'whale'/'fish'- and the 'arthritis' cases, I doubt that it can deal with all cases. In some cases of meaning change, our dispositions at the time of speaking to change usage upon counterfactually being apprised of information, give the wrong verdicts. Take the 'Jol'-case again. Suppose that if, counterfactually, Norway had been christianized in 800, and the vikings had been confronted by the religious authorities in the year 800, they would, *already then*, have agreed to change usage. The standard view predicts that as soon as the vikings acquired the deferential disposition towards the religious authorities, they meant the celebration of the birth of Jesus by 'Jol'. But is it true that the vikings already in the year 800 referred to the celebration of the birth of Jesus with 'Jol'? Intuitively, no. It seems false that the mere presence of such a disposition, non-actualized, in the year 800, entails that the vikings already then meant the celebration of the birth of Jesus. Furthermore, this would imply that a meaning change did not in fact occur in the

---

[33] It is fully consistent with changing usage that one does not regard one's previous usage as incorrect. This is tantamount to saying: "I used to use the word in a different way than I do now, and on that (previous) usage, my utterances were (mostly) correct". Labeling such dispositions to change usage a "correction" illicitly assumes that it is not a disposition to change usage.

[34] Something like the standard view does seem to be presupposed by traditional versions of semantic externalism, thought it is rarely (if ever) made explicit. As quoted above, Kripke says the following. "Scientific discoveries of species essence do not constitute a 'change of meaning'; the possibility of such discoveries was part of the original enterprise." (1980: 138) One interpretation of what Kripke calls the "original enterprise" is that speakers already before the essence-discovery had the overriding intention of using the given term as one whose reference is determined externalistically, that is, by the underlying structure or nature of the standard samples. These intentions can, further, be viewed as dispositions: It is, for instance, the fact speakers they *would* listen to scientists' findings, *would* look at things in the laboratory and not with the naked eye, and so on. More in detail: The reason speakers living before the nineteenth century did already then intend to use 'fish' in such a way as to exclude whales from its extension, and so that there was no meaning change, when the true biology of whales was discovered, is explained as follows. If these speakers had, counterfactually, say in the year 1600, been apprised of the true biology of whales, they *would have refrained from calling whales 'fish' already then*. The fact that what Art meant by 'arthritis' did not change upon being corrected by his doctor, but had always been diseases in the joints only, is explained by the fact that if he had been corrected, say, 5 years ago, he would, counterfactually, already then have been disposed to go along with his doctor's usage.

year 1000 after all, since 'Jol' already *then* did mean the Christian celebration, and that when the *actual* change in usage came about in 1000, there was in fact no meaning change. As a twist on the example, imagine that Norway were never in fact christianized at all, but continued to apply 'Jol' to the feast in honour of their gods for thousands of years. Imagine also that the vikings all the time had the disposition to change their usage of 'Jol' *should* Norway, counterfactually, be christianized at any point in history. Now, to hold that the mere presence of this non-actualized disposition entails that they all the time meant the celebration of Jesus by 'Jol' seems quite far fetched.

For these reasons, the standard view cannot be right as an account of meaning determination; the standard view takes meaning to be determined by speaker dispositions to apply a word at the time of speaking, under idealization. Since the vikings were already disposed to change the meaning of 'Jol' before the year 1000, the standard view is not able to capture the fact that 'Jol' changed only in the year 1000 and did not already in 800 mean what it did in 1000. Thus, even though the vikings' disposition to change meaning was already present long before the *actual* correction happened, it was clearly not meaning determining.

What has gone wrong for the standard view, in this case? It seems we must allow that even though the disposition is already present but unmanifested, it is still possible that *it is not meaning determining* for S′ uses at t_1. That is, the following principle seems true.

The Historical Contingency of Meaning What a speaker means by w at a time does not supervene on the speaker's total set of usage dispositions with w at a time, as meaning change can depend on whether speakers' dispositions to change usage are *actually* manifested. The counterfactual dispositions of two speakers at t_1 can be the same, yet they can mean different things by a term, depending on whether the dispositions are *actually manifested* or not.[35]

I will now argue that RDA, by contrast, gives the correct prediction about this case, as it is well placed to account for the Historical Contingency of Meaning. This is because it is solely concerned with *actual* usage, and not at all with counterfactual usage. According to RDA, we should look at the vikings' backward-looking corrective dispositions, from a future perspective, in which they possess all relevant information, to evaluate their past *actual* usage, and not simply at their present corrective dispositions. Presumably, the vikings did *not* think of their previous usage of 'Jol' as *incorrect* upon agreeing to change usage. They were not disposed to retract their earlier usage as incorrect, even though they agreed to change their present and future usage. If asked whether their earlier uses of 'Jol' were incorrect since they did not pick out the

---

[35] One important question to ask is why this thesis seems to hold only of meaning change, and not of correction of past usage. It seems there are two very different kinds of disposition in play. One is to be disposed to change usage because of what one sees as learning more information about *the phenomenon one was already concerned with*. This is what happens in discoveries of essence – one learns more about the nature of the phenomenon. The other is the one at work in meaning change cases; being disposed to change usage because of very different kinds of reasons: the reason for the disposition to change usage in such cases is not that one takes oneself to be learning more about the same phenomenon one already was concerned with. Rather, one takes oneself to be talking about something altogether new and different, not learning more about the same old phenomenon.

Christian celebration, they would most definitely *not* react by saying "We have all been wrong about what Jol was all this time", but rather "We used 'Jol' correctly before – to mean something different". Thus, RDA correctly predicts the past correctness conditions of 'Jol', that is, that the vikings' past usage of 'Jol' before the year 1000 was in fact correct, and so, meaning determining. Thus, it correctly predicts the meaning change in 1000; the meaning change took place precisely the moment the vikings *actually* agreed to change meaning, and not before this point. The standard view, by contrast, is not able to capture this verdict, at it appeals merely to dispositions to correct *at the time of speaking*, and as I have argued, these underdetermine whether a meaning change has happened.

Why is it that RDA seems to give the correct verdicts in the case? I argued that, at least in certain cases, it matters whether the dispositions to change usage are *actually* manifested or not. The standard view cannot capture this distinction, as it relies simply on dispositions, whether they are actually manifested or not. RDA, by contrast, is concerned solely with actual usage, and the correctness or incorrectness of this usage, not at all with counterfactual usage.

One might argue that if the only problem with the standard view is that it is not able to capture this, why not simply say that only actual manifestations of dispositions are meaning determining, and leave it at that? There are two problems. First, this fails to cover all cases. Although speakers' actual pattern of usage matters for meaning determination in the present case, it is not meaning determining in all cases. In some cases, counterfactual usage is enough to determine meaning. This is in cases where speakers correct their earlier usage, for instance in the 'whale'/'fish'-case. In these kinds of case, counterfactual usage is enough to determine meaning; it is the fact that speakers would (counterfactually) refrain from calling whales 'fish' that determines that whales are not in the extension of 'fish', even before the usage change actually happened. Second, the error-problem returns: it is not equipped to explain error, as it takes meaning to be determined simply by actual use. Indeed, the whole motivation for going for dispositional rather than actual usage, in the first place, was to solve the problem of error (as well as the problem of finitude) (Kripke 1982: 23–37).

Fortunately, RDA is able to answer both worries. Regarding the problem of error, RDA specifies that it is speakers' actual usage patterns that are meaning determining, but it does not simply *equate* meaning with such patterns. Instead, it provides an account of meaning determining conditions as those in which a speaker is, under the idealization of all relevant information, disposed to evaluate her previous actual usage as correct. Errors are explained as those cases in which a speaker is disposed to evaluate her earlier actual usage as incorrect. What RDA does, in effect, is to provide a higher order perspective on one's previous *actual* usage, and selects certain uses as correct and others as incorrect. Furthermore, RDA can give an account of speaker's correction of their earlier selves (the 'fish'/'whale'-case) just as well as the standard view. Instead of explaining the case by appeal to speakers' application dispositions at the time of speaking, we can look at their self-evaluative dispositions to evaluate actual usage, from a future perspective. The reason a speaker, living in the seventeenth century applied 'fish' uniformly falsely to whales, is that she is disposed, upon apprisal, to evaluate her earlier actual usage, as incorrect. The reason Art, already before his visit to his doctor, is to be ascribed the standard concept in his use of 'arthritis', is that he is disposed to correct his earlier self, from a future viewpoint.

Thus, as RDA is equipped to deal better than the standard view with meaning-change cases, and also can explain the correction cases as well as the standard view, we should accept RDA for both cases, in the interest of parsimony.

6 Further Development of RDA – Clarifications and Objections

In this final section, I will further develop RDA by first clarifying a number of issues pertaining to it, as well as answering some likely objections to it.

6.1 Do the Self-Evaluative Dispositions Have to Be Actually Manifested?

In a word, no. Although the *usage* that the self-evaluative dispositions react to has to be actual, the *dispositions to retract or to retain* a given response in light of all relevant information do not have to be actually manifested. RDA holds that it is enough that they are counterfactually manifested. The kinds of situations of appraisal of all relevant information are highly idealized and hypothetical; such appraisals do not often actually occur in real life. Most speakers will not actually find themselves in such a state of having all relevant information about their previous usage. However, in some cases they probably will. The claim of RDA is not that speakers do *actually* have such information and make such judgments; it is rather the hypothetical thesis that if the speaker *had* had such information, she would respond in such and such ways. This means that words can have referents other than those we believe them to have, if we are not in possession of "all relevant information".

6.2 Does RDA Make Meanings Inscrutable?

The foregoing means that if RDA is true, it follows that speakers will not have any kind of privileged epistemic access to facts about meaning, as they do not have the requisite access to the meaning determining facts, at least at the time of speaking. One might find this objectionable, as it seems, in our daily lives, that we do have access to knowledge about what our words mean. In response to this, it should be noted that many, or perhaps most metasemantic theories held today are not committed to such scrutability theses. For instance, traditional physical externalism holds that reference with proper names and natural kind terms is possible even in spite of ignorance and error. The same goes for social externalism. The currently popular view Reference Magnetism holds that meaning is determined by elite properties that our current science is probably ignorant of. Temporal externalism views meaning as dependent on future facts, to which speakers may not presently have any access, even in principle. So, if the non-scrutability of meaning is problematic for RDA, it is equally problematic for most contemporary views.

6.3 Temporal Externalism?

As I said in the previous section, most dispositional theories take meaning for a speaker at a time to be determined by her dispositions to apply w at that time under idealization; this is what I labeled the 'standard view' above. However, a different kind of view would be one according to which what a speaker currently means is not fully

determined by her present dispositions, but also partly by her *future* dispositions. Such views are instances of *temporal externalism*, according to which a speaker's *present* states are not enough to determine her words' meanings, but that *future* facts (actions, events, mental states) are also necessary (Jackman 1999). RDA is a form of temporal externalism; it takes speakers dispositions to apply a term at a given time of speaking, in itself, to be insufficient to determine its meaning at that time of speaking. In addition to these, speakers' *future* self-evaluative dispositions are also needed. RDA holds that the retraction dispositions of future time slices of a speaker play a constitutive role for what the speaker means now. The dispositions that RDA invokes have to occur at a future stage in history, since only in this way is it able to capture what I called the 'Historical Contingency of Meaning': the self-evaluative dispositions must necessarily be directed at *actual* usage through the history of a given term. To be placed to evaluate this usage, the dispositions must concern themselves solely with the actual usage, and they must necessarily operate from a future perspective as they need to survey the whole *actual history* of the term's use.

6.4 Individual or Communal?

One way to view the backward-looking corrective dispositions is to see them as strictly individualistic, such that it is the individual herself who corrects her previous applications. Another possibility is to construe the corrective dispositions as *communal*, such that it is the community of speakers that are disposed to correct individual speakers, in the idealized state of all relevant information. One might argue that if a theory of reference is to be consonant with semantic externalism (or even subsume it, as Johnson & Nado want to do), one should allow for correction to be communal, not only individual. If social externalism is true, individual speakers can count as competent with words through deference towards experts or the community at large. In that case, it seems that it is not necessary for individual speakers to be semantically competent that they actually operate under idealized conditions, as long as the speakers toward whom the individual is deferential operate under such conditions. Thus, if we opt for a communal version of RDA, it significantly lightens the semantic burden of individual speakers, although not for experts or for the community at large.

7 Conclusion

Johnson & Nado's theory, ReCIDA, is a promising newcomer in the field of dispositional metasemantics. It suggests a way to solve the problem of error that gives mostly correct predictions about meaning and reference. However, it is not without problems: in this article I have, first, identified some counterexamples to it. Second, I suggested a way to augment the theory to make it cope with one of the counterexamples. The counterexamples were, on the one hand, that ReCIDA cannot explain the phenomenon of attaining semantic competence as a result of gaining all relevant information. On the other hand, ReCIDA is unable to account for meaning change. I suggested that we can amend ReCIDA to enable it to deal with the latter counterexample. To this end, I outlined and defended a metasemantic theory, RDA, and argued that it can deal successfully with the second counterexample.

References

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

Blackburn, S. (1984). The individual strikes back. *Synthese, 58*(3), 281–301.

Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy, 4*(1), 73–122.

Burge, T. (1986). Intellectual norms and foundations of mind. *The Journal of Philosophy, 83*(12), 697–720.

Chalmers, D. (2012). *Constructing the world*. Oxford: Oxford University Press.

Cohnitz, D., & Haukioja, J. (2013). Meta-externalism Vs meta-Internalism in the study of reference. *Australasian Journal of Philosophy, 91*(3), 475–500.

Dickie, I. (2011). How proper names refer. *Proceedings of the Aristotelian Society, 111*(1pt1), 43–78.

Dretske, F. (1981). *Knowledge and the flow of information*. Cambridge, Massachusetts: MIT Press.

Evans, G. (1973). The causal theory of names. *Aristotelian Society Supplementary Volume, 47*(1), 187–225.

Fodor, J. (1987). *Psychosemantics*. Cambridge, Massachusetts: MIT Press.

Fodor, J. (1990). *A theory of content and other essays*. Cambridge, Massachusetts: MIT Press.

IAU. (n.d.). Definition of 'planet': https://www.iau.org/public/themes/pluto/

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.

Johnson, M., & Nado, J. (2017). Actual vs. counterfactual metasemantics: A reply to Andow. *Philosophia, 45*, 717–734.

Kripke, S. (1980). *Naming and necessity*. Cambridge, Massachusetts: Harvard University Press.

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

Laporte, J. (2004). *Natural kinds and conceptual change*. Cambridge: Cambridge University Press.

Millikan, R. (1984). *Language, thought and other biological categories*. Cambridge, Massachusetts: MIT Press.

Pettit, P. (1990). The reality of rule-following. *Mind, 99*(393), 1–21.

Pettit, P. (1996). *The common mind*. Oxford: Oxford University Press.

Putnam, H. (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science, 7*, 131–193.

Warren, J. (2018). Killing Kripkenstein's monster. *Noûs*. https://doi.org/10.1111/nous.12242.