Coordination, Content, and Conflation

Kyle Landrum

Coordination, Content, and Conflation

Kyle Landrum

Princeton University

ABSTRACT
Coordination is the presumption that distinct representations have the same referential content. Philosophers have discussed ways in which the presence of coordination might bear on the metasemantic determination of content. One test case for exploring the relationship between coordination and content is the phenomenon of conflation—the situation in which representations are about distinct things but are nevertheless coordinated. In this paper, I use observations about conflation to develop an anaphoric metasemantics for some representations in which coordination plays an integral role. I also develop some novel remarks on the problem of misrepresentation.

1. Introduction

Suppose that one day you watch a lovable borzoi bounding around a dog park that you frequent. You've never seen this particular dog before, but its playful elation warms your heart in a memorable way. The next day, you return to find a borzoi again springing about the park. You don't consider the possibility that the borzoi you're seeing is different from the borzoi you saw the day before: you *presume that they're the same dog*. Call this presumption of sameness *coordination*.

Philosophers have been interested in the semantic and epistemological relevance of coordination since at least the seminal Frege [1892]. One issue that philosophers have discussed is whether the presence or absence of coordination has any bearing on what the thought is about—that is, on the *content* of the thought. Two situations become salient, the first kind being the familiar *Frege cases*, in which two representations about the same thing fail to be coordinated. This would be your situation if you encountered the same borzoi but failed to recognize it. The other salient kind of situation focuses on how the *presence* of coordination relates to mental content. The cases of interest involve coordination between representations about distinct things. Call this kind of situation *conflation*. If it turned out that the borzoi you see today is different from the borzoi you saw yesterday, then you would be conflating the borzois.

In this paper, I will use some observations about conflation to argue for a principle relating coordination and content. The principle says that the presence of coordination is paramount in fixing the content of a particular class of representations. The class is determined by *aetiology*: whether coordination matters to the content of a

representation depends on why that representation formed in the first place. I'll also suggest that, for the concepts that don't have the aetiological profile of interest, coordination plays no role whatsoever in content determination, thus dividing concepts into a partition.

Why care about this principle? Because the principle has implications for our theorizing about mental content and the epistemology of knowing what we are thinking about. I will illustrate by applying the principle to an old problem—the problem of saying when you have *misrepresented* something. To take an example from Fodor [1990], suppose that you see a horse on a dark night and mistake it for a cow. The problem then arises, of how to answer the following: what makes it the case that you are indeed misrepresenting the horse as a cow, rather than correctly representing the horse as a cow-or-horse? While the latter class of entities is odd, it is nevertheless metaphysically innocuous, and could be talked and thought about (as we are doing right now). So why, exactly, do you count as misrepresenting cows rather than as representing the 'disjunctive' category?

I'll show that the principle developed in the paper's first part reduces the problem of misrepresentation to the problem of *original concept acquisition* for many cases and ensures that, in those cases, we cannot be mistaken about what we're thinking. In a more speculative vein, I'll also suggest that many cases of misrepresentation are cases of *one-off conflation*. I spend the denouement discussing the philosophical benefits of this reduction.

I begin with clarification of terminology and the key notion of coordination. I also state some of my background assumptions about mental content and mental representation.

2. Some Background Assumptions about Coordination and Representation

Our initial question is this: does the fact that one mental representation is coordinated with another have any bearing on the content that mental representation has? My answer will be that, for the mental representations formed in a specific way, it does. To begin spelling this out, we should start by getting clearer on the notion of a mental representation.

A mental representation is a kind of symbol, a particular on which cognitive processes operate. I'll focus primarily on propositional attitudes, which are one kind of complex representation. What makes attitudes complex is the fact that they have *constituency structure*—that is, an important part-whole structure. For example, my belief NINA SIMONE WROTE PASTEL BLUES is a complex representation partly composed of the representation NINA SIMONE.[1] The constituency structure of a representation is supposed to play a role in explaining what kinds of processes a representation enters into, and also why various constraints exist—for instance, why it's impossible to think THE LOUD DUCK without also thinking DUCK.

In what follows, I'll be assuming a familiar toy model of the attitudes, according to which they possess a constituency structure similar to the constituency structure of natural language sentences.[2] I will be concerned especially with the *bedrock*

---

[1] I am following the convention of referring to mental representations by using small capitals.
[2] See Rescorla [2019]. The contemporary *locus* point for the model is Fodor [1975].

constituents of attitudes, the atoms that don't possess any important constituency structure themselves. I will (with some trepidation) use the word 'concept' to denote these representational atoms, regardless of whether the constituent represents a particular or something more abstract like a category. So, my belief that NINA SIMONE WROTE PASTEL BLUES partially consists of the *concept* NINA SIMONE, which represents Nina Simone.

When I talk about attitudes and representations more generally, I am predominantly focused on *token* attitudes—the particulars that are located in a given brain and enter into various causal processes responsible for our individual behaviour. The aetiology of your representations and mine are usually distinct, even if those representations are type-identical: we each have token attitudes of the type VENUS IS THE SECOND PLANET FROM THE SUN, but the event of formation for the token that persists in your head is distinct from the event of formation for the token that persists in mine, as is the causal history of those respective tokens in our cognitive lives. Token representations can also cease to exist, because they are forgotten or because they are actively eliminated in some way. Some attitudes are relatively stable across time, like VENUS IS THE SECOND PLANET FROM THE SUN, while others are quite transient, like the thought THAT WAS AN ELEGANT HOP as deployed when witnessing the borzoi bounding about.

My focus on token attitudes and their token constituents necessitates an important cautionary note. In my terminology, there are two distinct token concepts in the belief VENUS IS THE SECOND PLANET FROM THE SUN AND THE ONLY PLANET IN THE SOLAR SYSTEM THAT ROTATES CLOCKWISE ON ITS AXIS IS VENUS, each of which falls under the concept type VENUS. But some theorists who work with the Mentalese model of thought would put this differently. They would say there is just one token concept that I possess—my token of the type VENUS—and that this token is deployed in two different *occurrences* when it constitutes those distinct attitudes. Their picture is that our token concepts are like words in our respective Mentalese lexicons, and that we reuse those Mentalese words to construct full-blooded (token) attitudes, in the same way that we have our natural language lexical inventories from which we select individual words when constructing sentences to utter.

We can translate the substantive points that I make below into this alternative way of speaking. Instead of asking how coordination of distinct tokens relates to the content of those tokens, we ask how coordination of distinct *occurrences* bears on the content of those occurrences. Notice that distinct occurrences of a token concept on the traditional way of speaking need not have the same content, even though they are occurrences of the same concept. This would be analogous to the way in which the word 'they' can have distinct contents in different occurrences of natural language discourse. So, we do not settle the questions of interest merely by rephrasing things in the idiom of occurrences. But I am putting things in my somewhat unconventional manner because the orthodox approach treats the coordination relation as tantamount to *reuse of a single lexical item*, and we naturally presume that uses of an expression like 'water' won't vary in content from context to context. I think that it's better to eliminate the potentially mistaken presupposition in the present discussion, since we are engaging with exactly the cases in which, arguably, the referential content shifts. If you find yourself struggling to count how many concepts are at issue in my sense, just remember to begin by identifying how many *attitudes* are at play and then to investigate their constituency structure.

Now I can say precisely what the *relata* of coordination are: they are token concepts. In particular, concepts are coordinated with each other just in case those concepts are presumed to have the same contents.[3] In my beliefs NINA SIMONE WROTE PASTEL BLUES and PASTEL BLUES HAS 9 SONGS, the distinct (token) concepts to which I referred via 'PASTEL BLUES' both pick out the very same LP. It is this object that I presume to be the same across the distinct thoughts.[4]

Coordination bears an intimate relationship with the inferential role of a concept. You believe that the borzoi was happy yesterday and that the borzoi is happy today. These two attitudes possess concepts that are coordinated. This coordination is closely related to your disposition to infer that the borzoi has been happy both yesterday and today, and also that there exists something that was happy on both days. Indeed, one way of characterizing coordination is as the relation that *rationalizes* such transitions in thought.[5]

This overview of coordination should suffice for my purposes. Now we can return to the initial project, but in precisified terminology: is there any important relationship between the fact that a token concept is coordinated with another concept and the fact that this concept has the content that it possesses? I want to convince you that, for each concept, it depends on the explanation for why that concept formed. The most straightforward way to argue for this is by discussing a case of conflation— that is, a case in which some concepts that are coordinated have distinct contents.[6]

---

[3] While the terminology is different, coordination orbits much of the deep discussion by Millikan [2000, 2017] of seeing things as the same ('same-seeing'). It is also closely related to 'trading on coreference' and 'internal co-reference' from Lawlor [2001, 2010] (the former is drawing on terminology from Campbell [1988]), 'the subjective appearance of coreference *de jure*' from Schroeter [2007, 2008], and the notion of a 'formal relation' from Heck [2012]. The word 'coordination' itself is used by Dickie [2015], Gray [2017], and Goodman and Gray [forthcoming]. Gray is drawing on Fine [2007] for whom the term 'coordination' is used to denote two different relations, one that involves the 'coordination' of the values of variables by a variable assignment in context and the other of which involves a kind of semantically necessary coreference, as illustrated by pronominative anaphora and reuse of the same name. Neither of these relations is what I'm talking about, although the second is closely connected to the relation that I'm discussing. In contrast, the relation discussed by Dickie, Gray, and Goodman is (I believe) the same relation that I have in mind. Also highly relevant is Pryor [2016, 2017]. The relation of coordination is closely related to notions like *cognitive significance* [Taschek 1995] and *mental filing* [Recanati 2012; Goodsell 2013]. I am not using this terminology because I think that 'coordination' makes salient the way in which my approach presupposes various facts about constituency structure, relations between constituents, and the primacy of token representations. These presuppositions seem less salient to me when using other terminology. Nevertheless, a lot of work on cognitive significance and mental files bears on what I have to say, and many of the authors just cited develop their theses in connection to those notions.

[4] The reader might be wondering about how definite descriptions fit into this picture, and, in particular, whether it makes sense to speak of concepts as being coordinated with complex mental analogues to descriptions, à la NINA SIMONE coordinated with the complex representation THE CREATOR OF PASTEL BLUES. My current view is that the relation between definite descriptions and their satisfiers is a special case of the relation between a function and its values, which is different from the relation of coordination, but nothing that I say below will depend on this issue. Relatedly, there might be something like coordination between distinct attitudes that does not depend on coordination between constituent concepts. An interesting example involving confusion about logical form is discussed by Ripley [2018: 172–3]. These examples also fall outside the scope of my argument.

[5] Gray [2017] and Goodman and Gray [forthcoming] both use the word 'coordination' to pick out the relation that rationalizes such transitions.

[6] For some work that has discussed the relationship between coordination and content, especially in the context of conflation, see Kvart [1989], Millikan [2000, 2017], Lawlor [2001, 2005, 2007], Camp [2002], Goldberg [2005], Schroeter [2007, 2008, 2012], Carey [2009], Dickie [2015], and Frost-Arnold and Beebe [2020]. These authors do not generally use the words 'conflation' and 'coordination'. Nor do all of them invoke a Mentalese model of the sort employed here. I should also say that the word 'conflation' itself is sometimes used for a slightly different but closely related notion. In this alternative sense, a *single token* concept conflates distinct contents, perhaps because they are equally good candidate contents. Conflation in my sense frequently engenders conflation in this other sense (see the discussion of the strong principle below).

3. Conflation in Slow-switch Cases

My primary task is to describe the concepts for which coordination is metasemantically important. I claimed that these concepts are distinguished by their aetiology. I'll now provide a pair of examples to motivate the aetiological distinction. The first example is an example where coordination does *not* seem to play an important metasemantic role. I think that beginning with such a case makes it easier to grasp the class for which coordination *does* seem to play such a role. Both examples use a well-known case of conflation as their skeleton, the *slow-switch* case from Burge [1998].

In a slow-switch case, the agent is moved to Twin Earth one night while asleep, and continues their life while oblivious to these interstellar shenanigans. In addition to encountering unfamiliar particulars from that point onwards (such as the Twin Earth version of your best friend), there can be doppelganger substances, kinds, and other sorts of phenomena. The classic slow-switch case involves XYZ,[7] a liquid that functionally resembles water but is chemically distinct—XYZ will suffice as a doppelganger for our purposes.

First, the case where coordination isn't important. Suppose that, after you wake on Twin Earth for the first time, you open the Twin Earth version of your favourite 'popular science' journal while you eat breakfast, and you find an answer column explaining why salt helps to melt frozen 'water' in the winter. The explanation involves salt's lowering the freezing point of XYZ. Nothing about this explanation makes reference to the chemical constituents of XYZ; it refers only to facts about salt, dissolution, and the thermodynamics of heat exchange at the boundary of the solid and liquid phases. Consequently, you don't read anything that would alert you to the fact that XYZ is different from water; indeed, the explanation is the very same as the explanation for why we use salt on terrestrial ice.

Now, there is a thought you form while reading this article, a thought that you would express by uttering 'Huh, salt lowers the freezing point of water.' Consider the concept corresponding to your use of 'water', and call this token concept C. C is about XYZ and not about water, even though C is coordinated with other concepts that are about water, and even though this article is your first representational encounter with XYZ. Moreover, if C is about XYZ despite its coordination with water concepts, it seems like C will be about XYZ regardless of how different your standing attitudes about water are. In particular, regardless of how much more (or less) you believe about water, C gets to be about XYZ because of the particular way in which it formed, via testimony. I will subsequently refer to C and concepts like it as *exogenous* (if you are unsure about whether C is really about XYZ, I have more to say to you below).

Now for a case where coordination seems paramount, a variant of the example of C: suppose that you actually were in a position to know how salt thaws ice, but that you had never reflected on this question. That is, your background knowledge of water and chemistry was sufficient to enable you to infer the mechanism underlying the salt *explanandum*, but you've never so inferred. Now, suppose that you never encounter the Twin Earth science article. Instead, you spend an uneventful first day on Twin Earth interacting unknowingly with XYZ: you have some glasses of XYZ, you take a

---

[7] In what follows, I treat 'XYZ' like a name for a substance, even though it presumably is supposed to have internal constituency structure in the way that the expression 'H_2O' does.

shower in XYZ, etc. You form various thoughts—and concepts—over the course of these XYZ-laden experiences. Then, prompted by how cold it looks outside, you start thinking about the chemistry behind salt and water's freezing point. You wind up with a belief that you would express by saying 'Huh, salt lowers the freezing point of water.' Crucially, also suppose that at no point in your reflections do you leverage any of the concepts that you formed in connection to your XYZ experiences. Call the concept corresponding to 'water' in your belief D. I maintain that D is about *water*, not about XYZ. This is so, *even if* you have formed some other attitudes about XYZ on that day and D is coordinated with those. I will be calling D and the class of concepts that it exemplifies *endogenous*.

What is the key difference between C and D? In the latter case, some of your background beliefs are playing a role in explaining why you formed the relevant thought and the token concepts by which this thought is constituted. In particular, the fact that D is *coordinated* with the constituents of the background beliefs is integral: it is partly *because* those concepts are being treated as about the same thing by the underlying mechanisms of inference that you are able to form D. The way that these inferential mechanisms operate is, in part, by being sensitive to the coordination profiles of our concepts. In contrast, to the extent that coordination plays any role in explaining why C formed, it is clearly not an exhaustive role; there is a crucial worldly input, in this case, the testimony of the article. This difference in the *explanatory* role of coordination yields a difference in the *metasemantic* role of coordination. That is the sense in which the distinction between exogenous and endogenous concepts is an aetiological distinction.

We can actually give a precise principle linking coordination and content for the endogenous cases. To get an initial sense for the principle, consider the following plausible claim about D: because D is produced in part via the coordination relation together with the mechanisms of inference, its content is simply *inherited* from the content of the water concepts that are explanatorily relevant to its formation—like a kind of mental anaphora.[8] That was why it was so obvious that D is about water, rather than XYZ.

More generally, concepts formed *because* of coordination relations—the endogenous concepts—seem to receive their contents from the concepts with which they are coordinated and that were responsible for their formation. Consider a generalization that you might form when encountering particular instances, such as the generalization that all swans are white. The concepts that constitute this generalization are formed, in part, because of the concepts constituting the generalization's inferential base and their coordination properties. Supposing that the concepts constituting these *base* beliefs all 'agree' in content—that is, that all of the concepts in the coordination profile for the 'swan' element of the generalization are about swans, and all of the concepts in the profile of the 'white' element are about whiteness—then how could the generalization fail to be about swans and whiteness?

---

[8] Burge [1993] discusses 'content preservation', and subsequent discussion can superficially appear to be about the same phenomenon that I am addressing: see, e.g., Lawlor [2002], who makes some suggestive remarks about 'mental anaphora'. Nevertheless, these theorists are primarily interested in something different from what I am focused on. In my terminology, they are interested in whether a token concept retains its content across different deployments in time—in other words, if content can change even while a token is being sustained across time in memory. The kind of content inheritance that I am discussing involves content inheritance across *distinct token concepts*, in virtue of the causal process that produced the tokens—in particular, one of inference.

Let me emphasize for a moment that it is only the coordinated concepts *that appear in the base beliefs* that are relevant to fixing the content of an endogenous concept; not *every* coordinated concept is relevant to that. The contents of coordinated concepts that played no part in the formation of a given token are metasemantically orthogonal. That was why it didn't matter that you had formed various XYZ-related concepts prior to your reflections about salt and ice. While it's true that D might be coordinated with a concept partially constituting a thought that you would express by saying 'I drank a glass of water this morning' (where we're assuming that this latter concept is about XYZ), that fact is irrelevant, because (by stipulation) that concept plays no role in explaining the formation of D.

This suggests the following principle.

> **Coordination and Content of Endogenous Concepts (Weak).** *For any (token) concept* R *formed endogenously, if every concept* R* *that is in the set of base beliefs causally responsible for the formation of* R *and that is coordinated with* R *is itself about some phenomenon* **o***, then* R *is about* **o***.*

What about the converse principle, which says that R is about some phenomenon **o** *only if* every coordinated concept in its base beliefs is about **o**? The key cases to investigate are again cases of conflation. In the case of D, the coordinated concepts in your base beliefs converge on a univocal content, but we can return to our original popular science article example, the one in which you form C. Suppose that the article nowhere mentions that the freezing point of XYZ is zero degrees Celsius. Concluding the article, you think something that you would express by saying 'So, since water freezes at zero degrees, it must be that salt lowers *its* freezing point below zero degrees.' Focus on the token concept that you articulate with 'its'. Call this concept F. We can see that F is endogenously formed: it is produced by reflection on two base beliefs, one of which is the belief that *water* freezes at zero degrees, and the other of which is the belief that salt lowers the freezing point of *XYZ*. We are already supposing that C, which is one of the concepts in the base beliefs of F, is coordinated with your water concept tokens, and so in particular it is coordinated with the token concept that partially constitutes your belief that water freezes at zero degrees. And it is partly *because* these concepts are coordinated that you form the thought you do, the one corresponding to 'salt lowers its freezing point below zero degrees.' The question is, what is the content of F?

It's clear that F has a special semantic status.[9] One plausible suggestion is that F has *multiple* contents: it is about both water and XYZ. If that suggestion were correct, then we could strengthen our principle to a biconditional and swap in an existential quantifier for the universal, like so:

> **Coordination and Content of Endogenous Concepts (Strong).** *For any (token) concept* R *formed endogenously,* R *is about some phenomenon* **o** *just in case and because* **some** *concept* R* *that is in the set of base beliefs causally responsible for the formation of* R *and that is coordinated with* R*, is itself about* **o***.*

While I believe that the strong principle is correct, I will only make use of the weaker principle. The important thing is that we've discovered a principle connecting

---

[9] For theorists who have discussed the semantics of cases like F, see Field [1974], Millikan [2000], Camp [2002], Lawlor [2005, 2007, 2010], Fine [2007], and Schroeter [2008]. Field is talking specifically about a phenomenon that he calls *indeterminacy*, which is closely related, but not identical, to conflation in my sense.

coordination and content for the endogenous cases, one that yields an anaphoric meta-semantics for these concepts. This metasemantics has interesting philosophical impli-cations, which I will develop momentarily in the context of the problem of error. However, before that, it's worth discussing the exogenous case a little more. In particu-lar, there are two questions on which to dwell. Is it really true that in the exogenous case coordination is playing no role in determining the content of the concept? Perhaps it plays a *partial* role. And even if it plays no role, *why* is that? The answer to this latter question could potentially elucidate why there is an exogenous/endogen-ous distinction at all.

So, first, why say that coordination plays no metasemantic role in the exogenous case? I observed in passing that C seems to be about XYZ regardless of how we vary the set of water concepts with which it is coordinated, and I concluded that coordi-nation doesn't affect the semantics of C or its exogenous brethren. But you might chal-lenge the claim that C is about XYZ from the outset, rather than about water (or, perhaps, about both). Indeed, you might challenge the claim exactly because C is coor-dinated with other concepts that are about water and because you take this to be meta-semantically important.

The first thing that I want to say is that, even if C is not about XYZ, I take it that there is still an important difference between it and D, and, in particular, that D is well-described by the principle connecting coordination and content for endogenous concepts, whereas C is not. If you agree that there is an important metasemantic con-trast here that is well captured by the principle, then I am satisfied and also expect you to find congenial the principle's application that will appear below. Nevertheless, let me offer some reasons for thinking that C really is just about XYZ.[10]

When I glossed the relationship between coordination and content in the endogen-ous case, I said that a (causal) *explanatory* relation between facts about coordination and facts about the formation of a given concept yielded a *metasemantic* relation. In general, philosophers have frequently called upon such causal explanatory relations (implicitly or explicitly) in making various content assignments. For example, consider the original Twin Earth case from Putnam [1975]: why do our pre-eighteenth-century thoughts not pick out XYZ, even if such a substance exists and satisfies equally well the properties that we associate with water? Because XYZ plays no role in explaining why our thoughts arose in the first place.[11] The slow-switch case raises new issues about mental content exactly because it involves moving you to Twin Earth, thus making XYZ explanatorily relevant to your representational practice.

That suggests the following: if we could show that water is *not* explanatorily relevant to the formation of C, then it would be ruled out as a candidate content, and XYZ would have no competition as the content. But, in order for water to be explanatorily irrelevant to the formation of C, *coordination* must be explanatorily irrelevant to the formation of C. Why? Because the only way that water could become explanatorily rel-evant to the formation of C is via facts about coordination. After all, it's certainly not

---

[10] Some theorists who offer a similar verdict are Gibbons [1996], Heal [1998], Goldberg [2005], and Schroeter [2007, 2008]. I'm not sure whether to include Burge [1988, 1998]. He claims that it takes some time to stand in the appropriate environmental relations to acquire a new concept—thus the nomenclature of 'slow-switch'—but does not say how long this is, and he is using 'concept' in a different sense than I am.
[11] Putnam was originally making an argument about *language*, and actually treated the twins as *psychologically identical*. So, this is an anachronistic way of putting his insight.

relevant via the testimonial path that runs from a fact about XYZ through to a fact about a sentence in Twin Earth English to a fact about you reading the article, etc.

So, is water explanatorily relevant to the formation of C? I confess that I'm pulled in both directions. But let me sketch the thought that would vindicate a verdict of irrelevancy. Distinguish between these two questions: why did some concept form *as opposed to nothing forming*, and why did some concept form *as opposed to another concept*? I think it's clear that, if we ask the first question in the case of C, water is completely irrelevant. To see why, consider a Good Case in which you're aware of your situation when you wake up on Twin Earth—the entities responsible for your superluminal relocation tell you what's going to happen before moving you. You don't take any of the particulars that you encounter to be the same objects from the day before, and you also know that you're no longer drinking water. Suppose that you begin your day in the same way as in our first scenario: you read a popular science article, while under no illusion that the article is about the terrestrial stuff with which you're familiar. You form a thought, a thought that you would express to the Twin Earth inhabitants by saying 'Salt lowers the freezing point of water.' Call the concept of interest deployed in this thought C^*. Clearly, C^* is about XYZ.[12]

Now, I take it that the answer to the question of why you formed some concept as opposed to not forming any is the same when we compare the cases of C and C^*. And because water is clearly not relevant in the case of C^*, it isn't relevant in the case of C either. All that's relevant is the causal pathway that runs from XYZ through the article through to the formation of your attitude.

If that's right, then we can glean why XYZ must be the content of C. More importantly, an appealing account of the exogenous/endogenous distinction emerges: it is grounded in a fundamental connection between facts about what is explanatorily relevant to the formation of a concept and facts about what the content of that concept is. For endogenous concepts, the things that are immediately explanatorily relevant are the coordinated concepts in the base beliefs that produced them. For exogenous concepts, these coordination facts are never explanatorily relevant.

In any case, my primary goal was to motivate the existence of the distinction itself and to articulate the anaphoric principle for the endogenous concepts, not to elucidate their foundations. I'll now show one philosophically satisfactory application of the principle.[13]

---

[12] If it's not clear at this point, I am presupposing relatively lax standards for concept possession: e.g. just acquiring a word for some phenomenon typically suffices for forming concepts of that phenomenon. For some important work that discusses conflation in the context of *richer* standards, see Camp [2002], Wilson [2006], Carey [2009], and Gupta [2011].

[13] It's worth briefly discussing Kitcher [1993], who argued that some uses of a theoretical term like 'dephlogisticated air' successfully referred to oxygen, while others referred to nothing. He invoked different metasemantic grounds for determining which uses worked and which ones failed. There is much overlap between Kitcher's view and the one developed here, especially in the idea that we should assign contents to tokens in context. However, he does not invoke anything like the exogenous/endogenous distinction: when trying to articulate the criteria for reference, he uses what he calls the 'principle of humanity' and a more interpretationist-friendly metasemantics. This makes content determination depend on facts about how it's best for us to interpret one another. While these facts might be extensionally adequate for figuring out the content of a concept, they are metaphysically removed from the facts about representational aetiology to which I appeal. He also frequently appeals to descriptions, on which I have nothing to say. Lastly, he is most interested in empty referent cases, about which I have said nothing. I believe that the picture that I develop needs to be complicated substantially to accommodate those cases.

4. Application(s) to the Problem of Error

Now we may return to the problem of error—the problem of saying why an application of a concept about *F* to some *G* that isn't *F* counts as a mistake rather than as the correct application of a concept with the content *F* ∨ *G*. There are many different extant strategies for solving this problem, including appealing to ideal epistemic conditions (what would you say if you had a better look at the *G*?), appealing to an asymmetrical dependence of the *F* ∨ *G* application to the *F* application, and, perhaps most popularly, appealing to some notion of phylogenetic or ontogenetic function.[14] But the foregoing discussion of coordination and conflation suggests a quite different way of thinking about the situation. In particular, it suggests breaking the problem into two parts, one corresponding to the cases of exogenously formed concepts, and the other corresponding to the endogenously formed ones.

Start with the latter class. Is there a problem of error when it comes to endogenously formed concepts? According to our principle relating coordination and content, endogenously formed concepts simply inherit the contents of the base concepts with which they are coordinated. In effect, then, we reduce the problem of error for endogenously formed concepts to the problem of *original concept acquisition*, or of how one was able to acquire a token concept that was about some phenomenon in the first place. Consequently, there is no real problem of error for the endogenously formed concepts.

I take this result to have some intuitive appeal. Reconsider the generalization that all swans are white. This generalization is false, because of the existence of black Australian swans. We could, in a sceptical spirit, raise the question of why we are allowed to say that this generalization is partially about the property of being white, rather than the property of being black or white. But a fair response at this point is to ask 'Where does the property of being black enter the picture?' It's not present in the causal story of how the generalization formed. And doesn't this story, together with the nature of inference, suffice for grounding the fact that the concept in the generalization is about the property of being white? We could wonder how the base concepts themselves 'glommed' onto the property of being white, but that is a separate problem. The naturalness of these ideas is probably part of the explanation for why theorists frequently do not discuss endogenous cases at all when discussing the problem of error: there is an implicit understanding that, if we can analyse the cases in which a concept *first* comes into contact with the phenomenon represented, we will have solved the hard problem.[15] That implicit understanding is vindicated by the principle that I've

---

[14] An exemplar of the first view is Strevens [2019], an exemplar of the second is Fodor [1990], and exemplars of the third strategy include Millikan [1984, 2000, 2017], Dretske [1988], Neander [2017], and Shea [2018].

[15] The reader might wonder whether the ruminations of Kripke [1982] bear on the endogenous cases. There, we are dealing with someone who computes a sum, which appears to be a paradigmatic case of inference. But Kripke's problem is that of determining what grounds the fact that the subject is thinking about addition rather than a different mathematical operation *at the outset* of this process. That is, when the subject considers a question like 'What does 68 plus 57 equal?', what makes it the case that the constituent corresponding to 'plus' is really about addition rather than some other similar operation? If we suppose that, at the beginning of the computation, they really do mean addition, then the output of this—the belief that 68 + 57 = 125—would certainly also be about addition, given the principle of coordination and content. But we would not have answered Kripke's sceptic, who is concerned with the starting point. In that respect, Kripke, like other theorists, is not interested in the endogenous case. Special issues arise in translating his problem into the toy model that we're using. One question is that of whether the content of a token representation remains the same while a token persists across time (see note 8). I have been taking for granted that it does, but one version of Kripke's sceptical problem is to ask why we should believe that.

given: for endogeneously formed concepts, the principle collapses the problem of their content determination to a question about how some given content was ever represented in the first place. I take it that this constitutes some progress.

The application also suggests that inference and other endogenous processes possess an epistemologically interesting property: the contents of your thoughts can't change over the course of an inference. If, at the end of some inferences, you find yourself thinking about something different than what you anticipated, it must be because you were conflating in your base beliefs. That is an attenuated but nonetheless substantive sense in which you could be immune to error when it comes to identifying the contents of your thoughts.[16]

So, if the anaphoric metasemantics is correct, there is no problem of misrepresentation for the endogenous concepts. What about the exogenous concepts? I'm going to close by considering a provocative claim. Recall the idea that coordination is wholly irrelevant to fixing the content of exogenous concepts. Now consider a situation like the Fodor scenario, in which you encounter a horse on a dark night but you think that it is a cow. Suppose that the 'cow' whinnies and you form a thought you would articulate by saying 'Wow! Cows can whinny!' I take it that the subject concept of this thought is exogenous. Consequently, the fact that it is coordinated with a bunch of cow concepts you have is metasemantically irrelevant. But if that's right, then what could the content of this concept be? If we take our verdicts about C in the slow-switch case and apply them here, then we're forced to say that you are actually thinking about *horses*. In general, mundane examples of misrecognition become one-off cases of conflation. The error that you make is not in forming a false belief about cows, but in forming a true belief about horses and then coordinating the concept in that belief with your cow concepts. On this picture, a case like Fodor's is, in effect, the least dramatic version of a slow-switch case.[17]

If you find this unappealing, you will no doubt question the claim that coordination is completely irrelevant to fixing the content of the exogenous concepts. But although this verdict about the exogenous concepts might be initially be unpalatable, there are benefits. Notice that again we reduce the problem of misrepresentation for a large class of concepts to the problem of *original concept acquisition*. To see why, note that in the Good Case where you recognize the horse as a horse, the concept you form satisfies some condition that determines *the property of being a horse* as its content. The story about exogenous concepts that we're considering says the very same thing about the Bad Case in which you take yourself to be encountering a cow. The condition that suffices for determining the property of being a horse in the Good Case also

---

[16] I thank an anonymous reviewer for encouraging me to emphasize this epistemological property.

[17] The idea that misrecognition should be construed as a different kind of error than plain false belief is something discussed extensively by Millikan [1994, 1997, 2000] (a good critical discussion of her evolving views is by Lawlor [2005]). The view from Millikan [2017] (which aims to be a comprehensive overview of her updated theory) is that misrecognition, at least in the cases that I'm discussing, is a different kind of thing from conflation: see her discussion [ibid.: ch. 6] contrasting false beliefs with 'equivocepts'. It's possible that the debate here is terminological; not all of Millikan's examples of equivocal concepts are clearly examples of conflation in my sense, and some seem to trade on her particular teleosemantic approach to the grounding of mental content (see her discussion [ibid.: 91] regarding Lorenz's goslings and Tweedledee and Tweedledum; only the latter case would involve conflation in my terms). In a very different vein, Camp [2002] argues that some of the humdrum cases that we would call cases of misrecognition should not be treated on the model of having a false belief. Indeed, Camp does not think that a semantic model of such a situation is appropriate at all, at least not in so far as the contents assigned are the familiar sorts of entities with which I've been working. For some of the reasons why I do not follow Camp's approach, see MacFarlane [2007].

suffices for determining the property of being a cow in the Bad Case. Of course, nothing that I've said explains why, exactly, the concept that you form in the Good Case is about horses rather than any other philosophical motley. But that is just the problem of explaining original concept acquisition, a problem that we need to solve anyway. So, subsuming the exogenous cases to one-off conflation in this way reduces the number of problems to be solved.

Additionally, nothing that I've said demands that we understand the *linguistic content* of your utterance 'Cows can whinny' to be about horses. That is, we could say that the *thought* that you would try to express by saying 'Cows can whinny' and the *sentence* you used might not have the same content. The linguistic predicate 'is a cow' would have the property of being a cow as its content (or, at least, nothing prevents us from saying that), in contrast to the token concept that you are trying to articulate by using that linguistic expression. Consequently, the sentence could be plain false even though your thought is true. That goes some way towards respecting any intuition that a plain false proposition is in the vicinity of your misrecognition.[18]

Allowing for a mismatch between thought and language content permits us to say plausible things about homonymous expressions. To see this, suppose that you tell someone on Twin Earth what you've read about salt and 'water'. In particular, you say something phonologically identical to 'Salt lowers the freezing point of water.' Intuitively, your audience can learn, from your statement, the mechanism by which salt prevents the freezing of XYZ: they can come to *know* something about XYZ from what you say. That would, on familiar approaches, typically involve you using some words, including a word about XYZ, to communicate a thought that you had about XYZ. The view developed here allows for this standard explanation, because it claims that C is about XYZ and it leaves open the possibility that you unknowingly acquired the Twin Earth word for XYZ when you read the article. We were able to say that C is about XYZ because it was an exogenously formed concept, as was the concept in Fodor's case. In other words, the same principle that entails the incongruence in content between the concept and the word in Fodor's example yields the intuitively correct verdict about how knowledge by testimony works in the slow-switch case. And we don't need anything as recherché as the slow-switch case to see that this phenomenon happens fairly often: I initially conflated Francis Bacon the painter with Francis Bacon the philosopher,[19] but, in spite of my confusion, I take it that my friend successfully learned that Bacon (the painter) painted the *Three Studies for a Figure at the Base of a Crucifixion* when I used the sentence 'Bacon painted the *Three Studies*.'[20]

---

[18] This mismatch between the content of your thought and the content of your sentence dovetails with older remarks about 'speaker's reference' and 'semantic reference'. Indeed, Kripke [1977: 263–4] uses a case of misrecognition in discussing the distinction. See also Kvart [1989].

[19] I'm embarrassed to say that seeing people wearing suits in his paintings did nothing to dispel my confusion, but only made me more astonished at how modern the paintings looked.

[20] To be clear, we can learn things from people who speak sentences that do not correspond to any concepts that they possess. Suppose that a Mandarin speaker is feeding me sentences over the phone, which I relay to you, but that I do not grasp, due to my inability to speak Mandarin (and suppose that you do speak Mandarin). Lastly, assume that all of us know that the speaker is feeding only truths to me. Then you can learn a lot of things even though I am merely parroting the sentences that I hear, without understanding them. Similarly, one could in principle learn true things about Bacon the painter from me even if I don't have a concept of Bacon the painter. However, there remain important differences between these cases and the conflation cases. Notice that if I am told that a sentence that I've just uttered is false in the phone case, I am at a loss as to what the upshot is: I have no grip on the significance of my having spoken falsely. If, however, I had reported an incorrect

Even if you still demur about the exogenous case, you can agree with me about the truth and philosophical import of the endogenous case and its associated principle. I'll close by briefly considering a parting worry: perhaps I've shown how to make progress on the problem of error by leveraging the endogenous/exogenous distinction, but doesn't a similar problem arise for the operative notion of *coordination*? That is, why think that the fact that two distinct token concepts are erroneously coordinated is any easier to understand philosophically than is the fact that a token concept has been misapplied to something that it doesn't represent? However, this question doesn't worry me. What we need to articulate in our explanation of coordination error are the grounds for coordination and the grounds for token concepts being about something from the start. As I said before, the latter grounds are something that we want regardless of what we say about (mis)recognition. But the former grounds are *also* something that we want, because coordination is an integral and independent facet of our psychology. And, crucially, I think that we can see that the grounds for coordination are amenable to naturalistic, and, in particular, computational analyses.[21] Unfortunately, rigorously spelling out such an analysis must be left for another time.[22]

Disclosure Statement

ORCID

Kyle Landrum http://orcid.org/0000-0001-7917-6401

References

Burge, Tyler 1988. Individualism and Self-Knowledge, *The Journal of Philosophy* 85/11: 649–63.
Burge, Tyler 1993. Content Preservation, *The Philosophical Review* 102/4: 457–88.
Burge, Tyler 1998. Memory and Self-Knowledge, in *Externalism and Self-Knowledge*, ed. Peter Ludlow and Norah Martin, Stanford: CLSI Publications: 351–70.
Camp, Joseph L. 2002. *Confusion: A Study in the Theory of Knowledge*, Cambridge, MA: Harvard University Press.
Campbell, John 1988. Is Sense Transparent? *Proceedings of the Aristotelian Society* 88/1: 273–92.
Carey, Susan 2009. *The Origin of Concepts*, New York: Oxford University Press.
Dickie, Imogen 2015. *Fixing Reference*, Oxford: Oxford University Press.
Dretske, Fred 1988. *Explaining Behavior: Reasons in a World of Causes*, Cambridge, MA: The MIT Press.

---

fact about Bacon the painter and you told me this, I am *not* at a loss as to what the upshot is. I will (on the story that I've told) know several things, such as that the person who painted *The Garden of Earthly Delights* is not Bacon. Thus, in the conflation case, if you wind up with a false belief, this entails that I must change my mind about something (to the extent that I am a diligent epistemic agent), whereas no such condition is in force when I am merely relaying Mandarin sentences to you. It's the difference between you learning something from me in my role as a knowledge *possessor* and learning something from me when I am a mere conduit for relaying truths. (Thanks to an anonymous reviewer for raising this issue.)

[21] I am sympathetic to *memory-address* approaches, which treat these dispositions as computationally realizing an *indirect-addressing* architecture. See Gallistel and King [2009], Green and Quilty-Dunn [2020], and Quilty-Dunn [2021].

[22] Thanks go to Eliya Cohen, Pavel Nitchovski, Chris Register, Laura Schroeter, Gabe Shapiro, Michael Strevens, and Snow Zhang for helpful conversations. Special thanks go to Grace Helton, Harvey Lederman, Gideon Rosen, Una Stojnić, and an anonymous referee from this journal.

Field, Hartry 1974. Theory Change and the Indeterminacy of Reference, *The Journal of Philosophy* 70/14: 462–81.

Fine, Kit 2007. *Semantic Relationism*, Malden, MA: Wiley-Blackwell.

Fodor, Jerry A. 1975. *The Language of Thought*, New York: Thomas Y. Crowell.

Fodor, Jerry A. 1990. *A Theory of Content and Other Essays*, Cambridge, MA: The MIT Press.

Frege, Gottlob 1892 (1997) (trans. Max Black). On *Sinn* and *Bedeutung*, in *The Frege Reader*, ed. Michael Beaney, Malden, MA: Blackwell.

Frost-Arnold, Greg and James R. Beebe 2020. Confused Terms in Ordinary Language, *Journal of Logic, Language and Information* 29/2: 197–219.

Gallistel, Charles R. and Adam Philip King 2009. *Memory and the Computational Brain: Why Cognitive Science will Transform Neuroscience*, Malden, MA: Wiley-Blackwell.

Gibbons, John 1996. Externalism and Knowledge of Content, *The Philosophical Review* 105/3: 287–310.

Goldberg, Sanford 2005. (Nonstandard) Lessons of World-Switching Cases, *Philosophia* 32/1: 93–129.

Goodman, Rachel and Aidan Gray forthcoming. Mental Filing, *Noûs*.

Goodsell, Thea 2013. Mental Files and their Identity Conditions, *Disputatio* 5/36: 177–190.

Gray, Aidan 2017. Relational Approaches to Frege's Puzzle, *Philosophy Compass* 12/10: e12429.

Green, Edwin J. and Jake Quilty-Dunn 2020. What Is an Object File? *The British Journal for the Philosophy of Science* 72/3: 665–99.

Gupta, Anil 2011. Meaning and Misconceptions, in his *Truth, Meaning, and Experience*, New York: Oxford University Press: 164–95.

Heal, Jane 1998. Externalism and Memory, *Aristotelian Society, Supp. Vol.* 72/1: 95–110.

Heck, Richard G. 2012. Solving Frege's Puzzle, *The Journal of Philosophy* 109/1–2: 132–74.

Kitcher, Philip 1993. *The Advancement of Science: Science without Legend, Objectivity without Illusions*, New York: Oxford University Press.

Kripke, Saul A. 1977. Speaker's Reference and Semantic Reference, *Midwest Studies in Philosophy*, 2/1: 255–76.

Kripke, Saul A. 1982. *Wittgenstein on Rules and Private Language: An Elementary Exposition*, Oxford: Basil Blackwell.

Kvart, Igal 1989. Divided Reference, *Midwest Studies in Philosophy* 14/1: 140–179.

Lawlor, Krista 2001. *New Thoughts about Old Things: Cognitive Policies as the Ground of Singular Concepts*, New York: Garland Publishing.

Lawlor, Krista 2002. Memory, Anaphora, and Content Preservation, *Philosophical Studies* 109/2: 97–119.

Lawlor, Krista 2005. Confused Thought and Modes of Presentation, *The Philosophical Quarterly* 55/218: 21–36.

Lawlor, Krista 2007. A Notional Worlds Approach to Confusion, *Mind and Language* 22/2: 150–72.

Lawlor, Krista 2010. Varieties of Coreference, *Philosophy and Phenomenological Research* 81/2: 485–95.

MacFarlane, John 2007. The Logic of Confusion, *Philosophy and Phenomenological Research* 74/3: 700–8.

Millikan, Ruth Garrett 1984. *Language, Thought, and Other Biological Categories: New Foundations for Realism*, Cambridge, MA: The MIT Press.

Millikan, Ruth Garrett 1994. On Unclear and Indistinct Ideas, *Philosophical Perspectives* 8: 75–100.

Millikan, Ruth Garrett 1997. Images of Identity: In Search of Modes of Presentation, *Mind* 106/423: 499–519.

Millikan, Ruth Garrett 2000. *On Clear and Confused Ideas: An Essay about Substance Concepts*, Cambridge: Cambridge University Press.

Millikan, Ruth Garrett 2017. *Beyond Concepts: Unicepts, Language, and Natural Information*, Oxford: Oxford University Press.

Neander, Karen 2017. *A Mark of the Mental: In Defense of Informational Teleosemantics*, Cambridge, MA: The MIT Press.

Pryor, James 2016. Mental Graphs, *Review of Philosophy and Psychology* 7/2: 309–41.

Pryor, James 2017. *De Jure* Codesignation, in *A Companion to the Philosophy of Language, Second Edition: Volume I*, ed. Bob Hale, Crispin Wright, and Alexander Miller, Malden, MA: Wiley Blackwell: 1033–79.

Putnam, Hilary 1975. The Meaning of 'Meaning', *Minnesota Studies in the Philosophy of Science* 7: 131–93.

Quilty-Dunn, Jake 2021. Polysemy and Thought: Toward a Generative Theory of Concepts, *Mind and Language* 36/1: 158–85.

Recanati, François 2012. *Mental Files*, Oxford: Oxford University Press.

Rescorla, Michael 2019. The Language of Thought Hypothesis, *The Stanford Encyclopedia of Philosophy (Summer 2019 Edition)*, ed. Edward N. Zalta, URL = https://plato.stanford.edu/entries/language-thought

Ripley, David 2018. Blurring: An Approach to Conflation, *Notre Dame Journal of Formal Logic* 59/2: 171–88.

Schroeter, Laura, 2007. Illusion of Transparency, *Australasian Journal of Philosophy* 85/4: 597–618.

Schroeter, Laura 2008. Why Be an Anti-Individualist? *Philosophy and Phenomenological Research* 77/1: 105–41.

Schroeter, Laura 2012. Bootstrapping Our Way to Samesaying, *Synthese* 189/1: 177–97.

Shea, Nicholas 2018. *Representation in Cognitive Science*, Oxford: Oxford University Press.

Strevens, Michael 2019. *Thinking Off Your Feet: How Empirical Psychology Vindicates Armchair Philosophy*, Cambridge, MA: Harvard University Press.

Taschek, William W. 1995. Belief, Substitution, and Logical Structure, *Noûs* 29/1: 71–95.

Wilson, Mark 2006. *Wandering Significance: An Essay on Conceptual Behavior*, New York: Oxford University Press.