Available online at www.sciencedirect.com
ScienceDirect
Cognitive Systems Research 9 (2008) 76–91
Cognitive Systems RESEARCH
www.elsevier.com/locate/cogsys

Frege's puzzle and Frege cases: Defending a quasi-syntactic solution

Action editor: Christian Onof

Robert D. Rupert

Department of Philosophy, University of Colorado at Boulder, Hellems 169, Campus Box 232, Boulder, CO 80309-0232, United States

Received 4 April 2007; accepted 1 July 2007
Available online 15 August 2007

Abstract

When a subject acquires a concept, one of her cognitive vehicles comes into an appropriate causal or informational relation to whatever that concept is a concept *of*. Social interaction helps in significant ways to ground this relation. I expound, then apply this perspective to a philosophical problem concerning conceptual content: Frege's puzzle. The socially interactive processes of language-learning and concept acquisition depend heavily on the mastery of reliable inferences involving the terms learned and concepts acquired. As a side effect, we are inclined to think that patterns of inferential relations constitute content itself. Thus, we are inclined to think that a subject's differing ways of thinking about the same object—i.e., her possession of two cognitive vehicles that refer to the same object but which participate in different patterns of subjectively drawn inferences—correspond to differences in mental, or conceptual, content. I argue that this is an illusion, an understandable one caused by the difficulty of language-learning and concept acquisition and the concomitant need to rely on inferential patterns to get ourselves into the appropriate causal and informational relations to the things represented by our thoughts and words. The illusion is strengthened, I claim, by the way in which subjects acquire the very concept of content.

Keywords: Mental content; Concept acquisition; Mental representation; Frege's puzzle; Frege cases; Symbol grounding; Language-learning

1. Introduction

There is no doubt that social interaction plays an important role in language-learning, as well as in concept acquisition. In surprising contrast, social interaction makes only passing appearance in our most promising naturalistic theories of content. This is particularly true in the case of mental content (e.g., Cummins, 1996; Dretske, 1981, 1988; Fodor, 1987, 1990a; Millikan, 1984); and insofar as linguistic content derives from mental content (Grice, 1957), social interaction seems missing from our best naturalistic theories of both.[1] In this paper, I explore the ways in which even the most individualistic of theories of mental content can, and should, accommodate social effects. I focus especially on the way in which inferential relations, including those that are socially taught, influence language-learning and concept acquisition. I argue that these factors affect the way subjects conceive of mental and linguistic content. Such effects have a dark side: the social and inferential processes in question give rise to misleading intuitions about content itself. They create the illusion that content and inferential relations are more deeply intertwined than they actually are. This illusion confounds an otherwise attractive solution to what is known as 'Frege's puzzle' (Salmon, 1986). I conclude that, once we have identified the source of these misleading intuitions, Frege's puzzle and related puzzles to do with psychological explanation appear much less puzzling.

2. Frege's puzzle

It seems fairly obvious that Mary—a randomly chosen, English-speaking adult with normal cognitive

E-mail address: robert.rupert@colorado.edu

[1] Both Putnam (1975) and Burge (1979) emphasize social factors, but neither offers a naturalistic account of mental content, where, by a 'naturalistic account', I mean one that appeals to no semantic properties, e.g., *referring to* or *being about*.

1389-0417/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogsys.2007.07.003

capacities—can believe that Mark Twain wrote *Adventures of Huckleberry Finn* (*AHF*) without also believing that Samuel Clemens wrote *AHF*. Imagine that, although Mary has basic knowledge of American literature, she does not know that Mark Twain's given name is 'Samuel Clemens'; in which case, it seems downright misleading to assert "Mary believes Samuel Clemens wrote *AHF*." Yet, a competing intuition must be given its due. There is only one person at issue, call him 'Twain-Clemens'. If Mary believes that *he* wrote *AHF*, then of course she believes that *he*, the very person Twain-Clemens, wrote *AHF*. Mary believes something about Twain-Clemens, regardless of how we refer to him when we ascribe a belief to her.

The first of these reactions to Frege's puzzle suggests that co-referring terms resist substitution in belief-contexts;[2] for such substitution might take us from a true sentence, "Mary believes that Mark Twain wrote *AHF*," to a false sentence, "Mary believes that Samuel Clemens wrote *AHF*." Frege (1892) himself concludes that belief-contexts are special, but his solution is not to deny the substitution of co-referring terms in belief-contexts. Rather, he denies that such terms co-refer when they appear in belief-contexts; instead, they shift their reference in these contexts, to refer to the what is normally their *sense*, the distinctive abstract concept or meaning that determines the reference of a term and that is grasped by every competent speaker who knows the term. It should be clear that the abstract meaning associated with 'Mark Twain' differs from that associated with 'Samuel Clemens'. Thus, Frege's view dissolves the puzzle. In belief-contexts, each name refers to a distinct sense; we substitute one for the other at our own risk, and understandably so.[3]

In what follows, I forgo the Fregean strategy and pursue the second of the intuitions bruited in connection with Frege's puzzle: that the content of Mary's belief has directly to do with the person, Twain-Clemens, regardless of how he is described in a belief-ascription. This is the Russellian view of belief content.[4] I do not argue directly for the Russellian view. Rather, I defend a Russellian diagnosis of Frege's puzzle and related problems of psychological explanation, thereby arguing indirectly for the Russellian view. This defense appeals partly to a theory of mental content; so let us set aside Frege's puzzle for the moment and think instead about the ultimate ground of the content of our mental representations.

3. Content-fixing and the symbol grounding problem

Consider the 'symbol grounding problem' (Harnad, 1990), the problem of attaching meaning to a mental symbol independently of the interpretation of theoreticians or modelers. Harnad presents this as a problem for a computational theory of cognition—the view that, roughly speaking, cognitive processes are species of computation as characterized by the mathematical theory of Turing.[5] The symbol grounding problem reflects, however, a more general challenge to any attempt to provide a materialist model of cognitive processing, as presently becomes clear. First, however, consider it as a challenge to computationalism.

The computational approach presupposes that at least some data structures are subject to semantic interpretation, i.e., at least some of the symbols in the domain of some of the relevant computations refer to, represent, or are about things in the environment (or other things in the computational system—let this be understood). For example, a computational account of chess-playing takes it for granted that at least some of the symbols in the domain of the relevant computations (e.g., those that generate various move-sequences to be considered) correspond to the various chess pieces and their positions on the board.

As an objection to computationalism, the symbol grounding problem takes aim at precisely this supposition: What, besides the modeler's stipulation, could make it the case that the symbols in question actually *do* represent chess pieces? Symbols appearing in a model refer to what they do simply because the theorists or engineers constructing the model say they do. In the graphic presentation of a model, the semantics of the set of primitive symbols might be encoded by a mnemonic or by presentation of an exhaustive list. This, however, is cheating if one's purpose is to give a naturalistic account of human cognition; the list or the mnemonic is effective only because those who read about the model understand the intentions of its creator. If the computationalist account of cognition presupposes an assignment of contents to the symbols, but that assignment itself requires the content-laden contribution of human cognitive processes, a blatant historical regress arises. When cognition first appeared, there was not anyone to assign meaning to the symbols in the domain of the relevant computations!

How should the symbol grounding problem be solved? To my mind, we should look to the philosophical theories of mental content developed in the last quarter of the twentieth century (Cummins, 1996; Dretske, 1981, 1988; Fodor,

[2] This talk of belief-contexts should be understood to apply to similar contexts, in particular, those involving other terms that express propositional attitudes, e.g., 'desire', 'hope', and 'fear'.

[3] On the Fregean account there is still some sense in which Mary thinks about the actual person, Twain-Clemens, but she does so only in an indirect way: part of the complex of customary senses to which she bears the belief relation—i.e., the customary sense of the name 'Mark Twain'—normally refers to the person, Twain-Clemens. Thus, a Fregean can hold that, in this derivative way, Mary's belief is about Twain-Clemens.

[4] I take a broad interpretation of the Russellian view, understanding it as a view about the contents of beliefs. On this way of dividing up the theoretical landscape, a view that assigns a causal or explanatory role to content-bearing vehicles (more on which below) qualifies as a Russellian view so long as the content of a belief is held to be nothing more than a construction from the worldly objects referred to and their properties.

[5] See Chalmers (1994) and Johnson-Laird (1983, chap. 1) for accounts of what it is to model cognitive processes computationally.

1987, 1990a; Grush, 1997; Millikan, 1984; Prinz, 2002; Rupert, 1999; Ryder, 2004; Stampe, 1977). According to many of these theories, causal or information-bearing relations provide the ultimate basis of mental content, and as such can be applied to computational symbols (or their realizations) in a straightforward way. [6]

Obviously, then, the literature offers a range of potential solutions to the symbol grounding problem (*pace* Gibbs, 2006, p. 159). I want to draw attention to two specific aspects of these theories that are relevant to my later discussion of Frege's puzzle. First, many causal-informational theories treat a mental representation's[7] content as being of (and perhaps in the first instance fixed by a relation to) concrete objects, individuals, or property-instances in the subject's natural environment. Such theories are meant to identify the conditions under which a given, mentally activated (or 'tokened') symbol correctly applies to some concrete object, state, instantiation of a property, or complex of such things. Some theories of this sort *also* explain how thoughts can be about general types, kinds, or abstract properties (Fodor, 1990a, chap. 4). Regardless, my point is that such theories should be attractive to anyone—a Russellian, for example—who takes mental content to consist primarily in a relation between the organism and individuals, kind members, or property instantiations in the environment.[8]

Second, consider the way in which causal-informational theories invoke, or can at least accommodate, causal mediation between a representational vehicle and the thing represented. Assume that the content of a symbol R in a given subject is a function of past causal interactions between tokenings of R and objects or property-instances in the subject's environment. Given that causation is typically (but not always) transitive, as is the relation *carrying-information-that*,[9] *R*s need not directly causally interact with the individual or instances of the property that R comes to represent. In the informational case, it is possible that *R*s carried information about, say, instances of P by having carried information about some state-type Q instances of which carried information about instantiations of P. Fodor, for example, emphasizes that the mechanisms mediating the content-grounding relation could even be other *beliefs*, so long as the content-fixing relation between R and what it represents can be characterized independently of the content of those mediating beliefs (Fodor, 1987, p. 121).[10] This second point is essential to the argument that follows: social and inferential factors can play an important mediating role in the fixation of content, without their being in any sense *constitutive* of the content so fixed.

4. Fodor and Aydede on Frege's puzzle

In his paper "Substitution Arguments and the Individuation of Beliefs," Jerry Fodor (1990b) presents a broadly Russellian solution to Frege's puzzle. Here it is, in short form. Let us assume the truth of an hypothesis for which there is a wealth of independent evidence: human cognitive processing takes place in a symbolic medium comprising mental particulars that play (or the realizers of which play) a causal role in cognitive processing (Fodor, 1975). According to Fodor, a subject believes that P, where P is the meaning or semantic value of her belief, if and only if she has a mental representation with the content P that plays the functional role of a belief. Something in addition to semantic value individuates belief-*states*, however. Fodor claims that the identity of the mental rep-

[6] We might wonder what evidence there is for such theories, beyond the intuitive appeal they hold for some philosophers. My view, which cannot be argued for in any detail here, is that cognitive science needs some account of the mind-world content relation, for reasons ranging from the large scale—the need to account for the truth of successful cognitive science's own claims—to the level of specific explanations. In this latter regard, imagine that Jill meets Jack on Tuesday, then recognizes him on the street the following Saturday. Presumably there is some portion of her cognitive system that was active when she met Jack and is active when she sees him later, and the activation of which on both occasions explains why she says "hello" to Jack on Saturday and not to any of the other of the hundreds of people she passes on the street on Saturday. If this mind-world relation plays an important role in cognitive scientific explanation—and in the explanation of what it is for a theory in cognitive science to describe anything accurately—then considerations of parsimony suggest that we try to do what we can with this one kind of content—the truth-conditional or, very broadly, the mind-world variety.

[7] I move freely between talk of symbols, mental representations, and terms in a language of thought. All such entities should be conceived of as having both semantic properties—they refer to or represent things—and syntactic properties, in the broad sense of 'syntactic' that entails only the existence of *some* way of typing such units that does not depend on their semantic values. I reserve the terms 'MOP' (short for 'mode of presentation') and 'vehicle' for referring to mental representations conceived of only syntactically. Sometimes I use such terms as 'mental representation' to talk about vehicles, in which cases I modify the term with 'conceived of syntactically' or 'individuated nonsemantically'. For more on the theoretical importance of this distinction, see Margolis (1998) and Rupert (1996, chap. 4, 1998, 2001).

[8] This is so even if some notion of a sense can be constructed from the materials of a causal theory of mental content (see, e.g., Devitt & Sterelny, 1987, p. 56).

[9] For the risk of nontransitivity in the causal relation, at least if that relation is construed as counterfactual dependence (Lewis, 1973), see McDermott (1995) and Hall (2000). In the case of information, a worry about transitivity arises in connection with Dretske's (1981) relativization of informational content to the state of the receiving system. If event a in system $S1$ carries the information that object o instantiates property$P$, and event b in system $S2$ carries the information that a is occurring in $S1$, one might expect that b carries the information that o has P. Nevertheless, b in $S2$ may fail to carry the information that o instantiates P, because $S2$ lacks a piece of additional information present in $S1$—with the result that $S2$'s information that a is occurring in $S1$ carries only the equivocal information that, say, o is in P or o is in Q.

[10] Although considerations of space preclude detailed treatment of the issue, note that causal-informational theories' allowance for mediation should assuage the anti-computationalist concerns expressed by Harnad (1990); for the mediating structures that help to provide symbols in a computational system with their content might consist partly of sensory routines, images, or templates of the sort emphasized by Harnad.

resentation—nonsemantically individuated—can make the difference. On this view, the content of a belief does not by itself individuate belief-states.[11]

Returning to our earlier illustration, Mary cannot believe that Twain wrote *AHF* without believing that Clemens wrote *AHF*. We do, however, distinguish between her state of believing that Twain wrote *AHF* and her state of believing Clemens did, allowing that she might be in the former without being in the latter; for she can have two different representational vehicles each of which represents Twain-Clemens, and given sufficient differences in the causal roles of these two different vehicles, their differential appearance entails a difference between the two relevant belief-states—as causally relevant psychological states—in Mary.

The basic idea behind Fodor's solution is straightforward and appealing. The contents of belief are Russellian. Attributions of beliefs, however, typically track belief-states. We want to describe subjects' beliefs in a way that reflects, as accurately as possible, differences in the causal powers of vehicles that are partly constitutive of belief-states. Thus, we describe Mary's beliefs *vis-à-vis* the author of *AHF* in a way that will capture the likely behavioral effects of the belief-state in question. We choose the name "Mark Twain" because doing so helps us to predict and explain her behavior more accurately—beginning with the simple prediction that Mary will express agreement when she hears someone utter "Mark Twain wrote *AHF*" but look perplexed or express disagreement when she hears someone say "Samuel Clemens wrote *AHF*."

Fodor's proposal may seem unproblematic applied to a single subject, but when we turn to intersubjective comparisons, difficulties arise (Aydede, 1998, 2000). If belief-states are individuated partly in terms of a mental representation conceived of nonsemantically, then any two people in the same belief-state must instantiate the same vehicle.[12] This, however, requires a way of individuating vehicles across subjects, no plausible version of which has yet been offered.

Furthermore, consider intersubjective patterns of behavior the explanation of which would, taking Fodor's approach, seem best explained by adverting to vehicles shared across subjects.[13] Assume that Superman is real. Imagine a group of people being approached by a powerful villain destroying everything in her path. Clark Kent stands prominently by the crowd, with accompanying TV cameras filming the scene (Kent has moved to television). Many members of the crowd scatter in fear. Many others gather round Kent, begging him to don his costume and fight the villain. Why? Intuitively speaking, members of the latter group believe that Kent is Superman, while members of the former group do not. Fodor seems committed to the following explanation: members of one group represent the person near the TV cameras, call him 'Super-Kent', via the activation of their MOP SUPERMAN, while others think of him via the activation of their MOP CLARK KENT. The group's divided behavior is explained by the members' differential tokening of these two different vehicles. This explanation, however, seems to require that members of each of the two subgroups share, among themselves, the vehicles SUPERMAN, on the one hand, and CLARK KENT, on the other; and Fodor offers no plausible way to individuate vehicles such that they can be shared across subjects, or so Aydede (1998) claims.

Let me emphasize the serious difficulties associated with the interpersonal typing of MOPs. One way to individuate mental representations nonsemantically appeals to their actual causal roles in the subject's cognitive system—their computational roles, for example. On this view, a MOP is individuated by the collection of causal interactions into which it is disposed to enter. In a particular subject, however, any given MOP is likely to be part of a complex network of the realizers of a large collection of beliefs, memories, etc., many of which are idiosyncratic (Fodor, 1990b, pp. 167–168). As a result, it is very unlikely that any two subjects share a MOP. An alternative endorses neurological criteria for the individuation of MOPs. This approach, too, seems to preclude the sharing of MOPs. Given the neurological facts—that the number of neurons and their pattern of connectivity varies greatly from subject to subject—it is highly unlikely that two subjects share functional neural structures. A third option suggests that the human cognitive system is best represented as a dynamical system (Port & van Gelder, 1995), in which case a MOP might be an attractor in the phase space of that system (Rupert, 1998). In a human cognitive system, though, the distribution of attractors depends to a significant extent on the subject's peculiar innate endowment and her actual past experience, and so, plausibly, will vary greatly from subject to subject.

5. Schneider on interpersonal Frege cases

In a recent paper, Susan Schneider (2005) argues that many of the concerns about interpersonal Frege cases can be addressed within a Russellian framework by adverting to facts about only *intra*personal stability in LOT terms.

[11] This will sound strange to many readers, but keep in mind the difference between belief-content as a relation that the subject stands in to propositions or other objects beyond the boundary of her skin and beliefs as concrete tokens *in* the subject that are supposed to account causally for her behavior. Given these two dimensions of belief, it seems useful to distinguish, even if only by stipulation, the content of the belief from the causally efficacious belief-state (cf. Block, 1986).

[12] Similar problems arise with respect to our choice of words in belief-attribution. If, when attributing beliefs, the choice of a particular term ('Mark Twain', for example) is supposed to track causal powers, then, since the specific usage in question is meant to apply to a substantial population, it would seem that members of the population had better share vehicles with the same causal roles for that term to track.

[13] Questions concerning the best way to explain patterns of behavior take us beyond Frege's puzzle proper, to questions about the causal-explanatory work beliefs and their contents are meant to do (and to questions about what beliefs and their contents must be like in order to do that work). Here, we encounter a host of puzzles closely related to Frege's, many of which are commonly referred to as 'Frege cases', more on which below.

Here she draws from the work of David Braun (2001), borrowing, in particular, his talk of "matching ways." Imagine a person in frightened crowd who sees Clark Kent by the camera person and, at the same time, wishes Superman were present, yet fails to ask Kent for assistance (or direct any other superhero-appropriate behavior toward Kent). From the standpoint of a purely Russellian psychological theory (i.e., one that takes only the Russellian content of propositional attitudes to do causal-explanatory work in psychology), this person's behavior seems odd. Someone who wants Superman's help and who believes Superman is present should, other things being equal, ask for Superman's help. Our crowd member does not, however, because her various mental states ("I wish Superman were here" and "There's Clark Kent") represent Super-Kent in mismatching ways. The *contents* of the relevant mental states, when co-present, suggest one form of behavior, but an accounting of the vehicles in play explains why our hypothetical crowd member does not behave in a superhero-directed fashion.

Schneider offers the following general principle:

> (FP) *Ceteris paribus*, if system S has distinct MOPs that represent entity *a*, but the MOPs are not linked in the system's data set as being coreferential, and S's behavioral success depends on the system's linking this data, then: S will appear to act irrationally in her *a* directed action. (Schneider, 2005, p. 438)

Part of Schneider's brief is to show that Russellian psychology can treat Frege cases as cases where other things are *not* equal relative to the laws of Russellian psychology itself.[14] Of greater importance for my purposes is the intrapersonal nature of the two LOT terms being treated, or not being treated, as coreferential. Whether the system treats two terms as coreferential would seem to be a fact about only the way in which *that* system processes the two terms, regardless of whether any other system instantiates type-identical LOT terms. As Schneider puts it, "[D]ifferent systems may satisfy (FP), despite the holism of computational state individuation, insofar as each system has two (intrapersonally distinct) MOPs that represent the same entity" (Schneider, 2005, p. 445, n33).

If Schneider is correct, the lack of intersubjectively shared MOPs, emphasized above at the end of Section 4, is a red herring. Each member of the frightened crowd has two MOPs that refer to Super-Kent, but they need not be the *same* two MOPs across subjects. This view does not presuppose that there is on the table a clear and convincing criterion for individuating MOPs. Rather, Schneider need claim only that we can explain the shared behavior that worries Aydede without assuming shared MOPs across subjects. This requires only that each subject possess some pair or other of MOPs that exhibit a certain higher-order relation (matching or mismatching). So long as we are confident that, in the individual subject, some structures or other, individuated nonsemantically, bear semantic relations to the environment and affect behavior, this approach can do without any specific theory of how MOPs are to be compared across subjects.[15]

I will propose a variant on Fodor's and Schneider's views, but first, consider a complication in the way the

[14] Schneider is here pursuing a tack taken by Fodor (1994). Complex issues arise in connection with *ceteris-paribus* clauses, which issues I cannot treat in any detail here. Consider, though, one independently motivated consideration that speaks fairly directly to the issue. Assume that psychology is intentional in Fodor's sense: psychology couches its generalizations in terms of the Russellian intentional contents of the psychological states involved. *Computational* modeling enters this picture as an explanation of the way in which intentional laws or generalizations are implemented: syntactically described processes can model the important semantic relations preserved in human thought (the ability to perform valid, deductive inference is the paradigmatic case—see Fodor, 1994, for a description and endorsement of this reading of computationalism's contribution to psychological theory). But, when a subject represents the same object via two different vehicles and is not disposed to treat those mental representations as referring to the same object, the appropriate semantic relations are not satisfied: the subject will fail to make certain obvious valid inferences. The kind of irrationality observed in Frege cases is thus outside the scope of what the computational theory is intended to explain *as a theory of the implementation of Russellian intentional psychology*, and is thus a case where other things are not equal from the standpoint of Russellian, intentional psychology. All the same, this approach leaves plenty of room for computational explanations of bad inference, faulty memory, etc., but again, such explanations fall outside the scope of computational modeling as a theory of the implementation of intentional psychology. Rather, these computational explanations would be offered at a lower-level than the level of content. Bad inference, faulty memory, etc., are likely to be explained as breakdowns of, or idiosyncrasies in, lower-level mechanisms, which do not have anything particularly to do with intentional *content*, and thus, do not seem amenable to either a Russellian or a Fregean solution.

[15] One reader wondered whether the view on offer makes sense absent an articulated theory of what it would be for subjects to share MOPs. This is tricky. It is true that Fodor's and Schneider's views (and the variant I develop below) presuppose that MOPs are well-individuated within an individual; if questions about MOP-identity within a single subject have no determinate answers, MOPs' alleged causal contributions to the behavior of even a single subject become suspect. Nevertheless, no one pursuing the kind of approach discussed in the text need commit to any particular theory of MOP-individuation in order for the proposal to make sense, so long as there are some extant candidates (which there are—see Section 4) that express fairly precise identity conditions for MOPs within a single subject. Furthermore, the kind of view on offer need not hold specifically that there are or are not shared MOPs. It might be that the best theory of MOP individuation does, after all, allow shared MOPs, in which case the pairs of MOPs that bear a common contrast across subjects could, in fact, be MOPs of the same type. All the same, the correct theory of MOP-individuation within an individual may not yield any shared MOPs in the case of humans. That would be acceptable as well and, notice, would not entail that cross-subject comparisons are incoherent—only that the comparison always yields the same result, *viz*. "not shared." Finally, note that even if our best theory of MOP-individuation within an individual does not permit coherent comparison across subjects, the kind of view on offer is unscathed. If a MOP in one subject cannot be coherently compared to a MOP in another subject, then those MOPs are not of the same type, i.e., they are not shared. This result does not prevent pairs from exhibiting the same higher-order contrast, so long as that contrast can be spelled out in way that does not presuppose the sharing of the MOPs themselves.

stage has been set. Following Schneider, I have spoken as if there is no problem of intrapersonal individuation. Some of the concerns raised above about interpersonal individuation of MOPs might, however, extend to the case of a single subject. The causal dispositions of what we think is the same persisting concept change over time within a single subject, as do the components of at least some of the relevant groupings of neurons, as do the shapes of the phase spaces of subjects' cognitive systems considered as dynamical systems. Why, then, should we think that within the individual, there will be any useful sense in which a given MOP might reappear in various contexts?

I offer two comments in response. Firstly, it is not clear that the story I am about to tell requires anything more than the one-off appearance of MOPs, although if one-off MOPs are the only (or the primary) kind appearing in human cognitive systems, I would have to reformulate my view significantly. Secondly, a more promising strategy appeals to the causal history of the set of MOPs within an individual as part of their individuative criteria, e.g., by appealing to relations of causal ancestry among the various MOPs existing in the individual at various times (Rupert, 1998). For example, in a single organism, there may be a gradual and orderly causal process by which one cognitively relevant set of co-firing neurons loses some of its cohort, while also recruiting new members. A relation of causal descending thus can provide a principled basis for the claim that, from the standpoint of a single cognitive system, there is one MOP persisting over time.

Let us move, then, to the quasi-syntactic treatment of intersubjective Frege cases. Understood in the broadest terms, Aydede's example includes two relevant groups: crowd members who beg the bespectacled TV reporter for help and those who do not. We should not dwell long on the first case, but note the complications introduced by subjects' mental representations of the *words* used to attribute beliefs. Each subject in the know might use a single vehicle to represent Super-Kent (although, again, different vehicles for different subjects) that is connected to two others, one representing the word form 'Superman' and another representing the word form 'Clark Kent'. Such elaborated structure is independently motivated; even subjects in the know tend to say "'Kent' has four letters" but not "'Superman' has four letters." Furthermore, many subjects in the know maintain the secret. This can be explained by invoking representations of word forms in addition to single MOPs representing Super-Kent. Such subjects might have beliefs of the sort "if someone talks about Super-Kent and uses the word form 'Clark Kent', then, absent some reason to think that person is in the know, do not talk as if 'Clark Kent' and 'Superman' refer to the same person."

What about the case of crowd members who all run away? Here the appeal to contrasting vehicles—mismatching ways—is indispensable (Schneider, 2005, p. 434). Take a crowd member who, while looking at Super-Kent dressed as a reporter, tokens one vehicle representing Super-Kent and who tokens a distinct vehicle while wishing for Super-Kent's aid (if that belief is activated at all); such a crowd member does not approach Super-Kent for aid. This is understandable if, at the nonsemantic level—the level of computational processing, in particular—the subject does not treat the two vehicles in question as co-referential. If a number of crowd members possess vehicles exhibiting the same contrast, we begin to see why they might all fail to ask Super-Kent for aid. Thus, something common to the group members—a shared contrast in the treatment of two of their vehicles, and perhaps the tokening of only one of them on a particular occasion—explains the similarity in the behavior at issue, even though these people may not share the MOPs in question.

Consider the following, more detailed story about such crowd members.[16] Each has a belief $b1$ with the content *Super-Kent is powerful* and a belief $b2$ with the content *Super-Kent is wimpy*. $B1$ and $b2$ are realized (or implemented) by some nonsemantically characterizable sentences in her LOT, call them $s1$ and $s2$, which contain as components vehicles $t1$ and $t2$, both of which refer to Super-Kent; furthermore, when $b1$ becomes occurrent, $s1$ realizes $b1$ and $t1$ occupies the syntactic position of subject in $s1$, and when $b2$ becomes occurrent, $s2$ realizes $b2$ and $t2$ occupies the syntactic position of a subject in $s2$; and $t1$ is a distinct from $t2$.[17] Typically, the subject's tokening of any LOT sentence with the content *that guy has a red cape and blue tights* activates $t1$; and, typically, the subject's tokening of a belief with content *that guy wears spectacles and appears on the nightly news* activates $t2$ (where activation can be cashed by co-opting whatever neural mechanisms account for priming effects). More generally, MOP $t2$ is causally connected to other MOPs and appears as a constituent of various LOT sentences. Some of these other MOPs, when activated, cause the activation of $t2$: such MOPs would likely include those carrying the contents *being a reporter* and *being bespectacled*, which themselves can be activated by perceptual processing (e.g., by the process of the seeing of Super-Kent dressed and functioning as a reporter). Other MOPs are causally connected to $t2$ in a converse manner: the activation of $t2$ causes their activation. These are likely to carry such contents as *being mild-mannered* and *lacking self-confidence*; but as indicated in the cases of $b1$ and $b2$, some of the phenomena of interest involve the tokening of sentences in LOT of which $t2$ is itself a constituent.[18] We then explain various subjects' running away, for example, by citing the activation of $t2$ rather

[16] Although I take Schneider's (FP) and some of Fodor's remarks as points of departure, I do not intend what follows to be a faithful exegesis of either's view.

[17] Bear in mind that the terms $t1$, $t2$, $s1$, and $s2$ are subject-specific. That is, for each subject, there exists some pair of terms or other (which for each subject I re-label $t1$ and $t2$) and some pair of LOT sentences or other that play the role of $s1$ and $s2$. Thanks to Murat Aydede for suggesting ways to formulate the ideas in this paragraph.

[18] Talk about the contents of these other MOPs to which, e.g., $t1$ and $t2$ are connected is sometimes put in terms of differing file folders or dossiers containing different collections of information (see, e.g., Forbes, 1989).

than *t*1, the contents of the beliefs or perceptual states that caused the activation of *t*2, and the contents of the further mental states either the realizers of which include *t*2 as constituent or are partly caused by *t*2 to become active. For instance, we advert to beliefs about, say, the dangerousness of super-villains and the prudence of running when there is no one to protect you from an attacking super-villain. If *t*1 is active at all, it is only as part of a desire that, by itself, does not control any of the relevant behavior (let this be understood hereafter).

To be clear, three factors explain shared behavior of the crowd-members who are not in the know (and here the behavior I have in mind is both their running away and their not asking Super-Kent for help). First, each subject of interest possesses a pair of distinct MOPs both representing Super-Kent. Second, in each subject of interest, only one of these two relevant MOPs is active. Third, for each subject, the active MOP fits a certain description, one given in terms of the *contents* of other MOPs (or strings of them) to which the active MOP is causally connected. Inclusion of this third factor renders my proposal only *quasi*-syntactic: while not necessarily sharing *any* MOPs, each subject has some MOP or other that stands in important relations to states individuated by their content. Thus, to explain behavior in Frege cases, we appeal partly to syntactic and partly to semantic facts: each subject in question has two different MOPs referring to the same thing; this difference explains why someone who has the standing beliefs that Super-Kent is powerful and good and that Super-Kent is standing nearby nevertheless fails to ask Super-Kent for help in the face of danger (take the belief that Super-Kent is powerful and good to play the role of *b*1 above). In this way, a syntactic fact plays an important explanatory role. Nonetheless, in order to explain why many such subjects exhibit shared behavior, we must also characterize the shared relations into which (possibly) subject-specific MOPs enter, and we do so in terms of the indirect causal relations that (possibly) subject-specific MOPs bear to mental or cognitive states the contents of which *are* shared across subjects. Much is to be explained by the fact that, of a subject's two MOPs, one is active and causally connected to MOPs (or appears as a constituent of a string of MOPs) that partly realize various other mental states.[19] Furthermore, this pattern of connections can be shared across subjects because the pattern is described in terms of external content and so can explain shared behavior in cases where shared behavior occurs and, in fact, should be given a unified explanation.[20] Notice further that the preceding approach can ground an explanation both of specific, shared behavior (e.g., running away) and of the simple fact that various subjects exhibit the irrationality associated with Frege cases.

What, though, actually accounts for the running from our super-villain? Focus on the network of other mental states to which the MOP that *is* tokened is related (by *being* part of, or by being causally related to some of, those mental states' possibly subject-specific realizers). Each subject in question tokens a particular MOP that represents Super-Kent. The tokening of this Super-Kent MOP causes the tokening of certain other MOPs (or contributes to the formation of certain LOT strings of which it is a part). Given the content of these other MOPs, for example, *mild-mannered-ness*, and the resulting content of the LOT sentences of which they are a part, we can subsume the subjects' subsequent behavior under generalizations put in terms of the Russellian contents, e.g., the psychological generalization "other things being equal, if one person believes another person to be mild-mannered, weak, and lacking in self-confidence, then the first person will, if possible, flee when faced by a powerful attacker, rather than asking the second to fight off the attacker."[21]

Clearly, which MOP is tokened makes a causal difference. Given that speakers are normally interested in tracking the causes of behavior, and given that our word choices in fact track some of these differences, speakers resist free substitution of co-referring terms when attributing beliefs

[19] It might appear that I am invoking inferential roles to individuate MOPs or to help to explain behavior. Given, however, that the "inferences" are cast partly in pure-causal terms, these are not exactly inferential roles; inference is often taken to be a semantic notion. Central to my explanation is the claim that *t*2's tokening causes the activation of some further MOP. Note also that the quasi-inferential roles in question are not assumed to individuate any MOPs or to constitute the content of any MOPs.

[20] Jerry Fodor pressed me to explain how such a view can avoid appealing to some kind of holism of the relevant inferential networks. Assume that a psychological explanation of the shared behavior appeals both to a shared network of beliefs, considered in terms of their content, and to the beliefs' realizers, which are causally connected to the subject-specific versions of *t*2. If this is correct, it might be thought that my proposal faces a serious problem: I need a way to individuate the network that must be shared in order for the various subjects to be subsumed by the generalizations in play. Networks would appear to be the sorts of things that are only partially overlapping and are subject to holistic principles of individuation; thus, they can no more be shared than can, say, neurally individuated MOPs. I do not think there is a problem here. Each of the inferential relations I have in mind can be characterized in terms of causal relations between beliefs with externalist content—no holism here. Furthermore, to the extent that various subjects' behavior *should be* subsumed under the same generalization, the behavior issues from the same externalistically individuated beliefs (i.e., beliefs thought of in terms of their Russellian contents). So, I have not presupposed the existence of a definite thing, a network, that all of the subjects share. Rather, in cases where the same explanation applies to a group of subjects, those subjects share some beliefs—but only the ones causally responsible for the behavior. On a realist view of these states and their causal powers and an externalist theory of content, subjects can share some subset of their various beliefs, without this entailing holistic individuation of networks; furthermore, precisely those shared beliefs can be causally responsible for the similarities of interest in the subjects' behavior.

[21] Note further that the same behavior can be caused by different thought processes or practical syllogisms, so there need be no entirely unified explanation of the behavior of the members of the crowd who flee and ignore Super-Kent.

to others. Free substitution within belief-contexts would cause predictive failures, because, for example, the word form 'Samuel Clemens' normally tracks the effects of whatever goes on in a person's mind specifically related to little-known facts and unknown persons; these goings-on do not produce such behavior as running towards Twain-Clemens and asking for an autograph. Of course, the average speaker does not think about MOPs and details of cognitive processing, but she can nevertheless be confident that some attributions have more predictive and explanatory power than others.[22] Our use of word-forms in belief-attribution is meant to track the causal relations that cognitive vehicles enter into, but not under that description; for we do not explicitly represent detailed facts about those vehicles. Instead, our tracking of the important causal differences is typically mediated by a lot of folk wisdom concerning the conditions under which one should say, e.g., "Mary believes Samuel Clemens wrote *AHF*." For instance, one should say *that* only if one has reason to believe that Mary has had certain experiences with the word form 'Samuel Clemens'—she has heard someone say, "Samuel Clemens *is* Mark Twain"—or if, given the context, truth-conditions are all that matters.

Does the approach I have recommended cover the full range of phenomena related to Frege's puzzle? Consider two related questions about belief attribution and intentional psychology. First, ask whether the perspective articulated above can account for the following true generalization:

(G) Those who believe that Superman is present feel safer.

Yes, so long as we understand (G) to apply positively (as opposed to vacuously) only to those who meet the following two conditions:

1. The subject has some MOP or other, $t1$, referring to Super-Kent.
2. MOP $t1$ is activated when the subject's MOPs with the content *wears a red cape*, *can fly*, etc. are activated. The truth of (G) then amounts to the truth of a further claim:
3. MOP $t1$, when activated, causes the activation a string of MOPs with subject-predicate form such that the MOP in subject-position refers to the subject's self and the MOP in the predicate-position refers to the property of being safe.

Of course, (G) is false if condition 2. is not included, for a subject might have some MOP or other, $t2$, with the semantic property of referring to Super-Kent but the activation of which is not caused by MOPs that represent red-capes, the ability to fly, etc. We do not expect this subject to satisfy condition 3. if $t2$ is the only MOP representing Super-Kent that is active while she is in the presence of Super-Kent. Such subjects, however, are people about whom we would resist saying, "She believes Superman is present." We would resist it precisely because, even though the content of her occurrent belief might be a Russellian proposition *that Super-Kent is present*, the choice of the word 'Superman' tracks the causal powers of a nonactive MOP.[23]

Now consider a second phenomenon. It seems obvious that two people can share the concept of Superman without sharing the concept of Clark Kent. If, however, concepts are individuated only by their externalist content, then anyone who has a concept of Superman also has a concept of Clark Kent. So, the question arises how, on my view, any subject can have one concept without the other. In fact, this is the primary puzzle. If this puzzle can be solved for the individual subject without appealing to specific individuating conditions of the MOPs involved (e.g., that the MOP involves such-and-such specific neurons), then an explanation of sharing is straightforward: the explanation of the individual case is true of all people in the relevant group.

How does the quasi-syntactic view handle the individual case? First, distinguish concepts from *conceptions*. Concepts are, in the first instance, atomic structures; conceptions are always compound. A concept can represent a single individual, property, or kind simply in virtue of a causal relation to the thing represented, whereas a conception includes beliefs or theories about an individual, property, or kind (cf.Cummins's discussion of knowledge structures—1996, pp. 132–135). Granted, I am initially inclined to think that someone could have the concept of Superman without having the concept of Clark Kent. Nevertheless, my intuition is mis-described. I am confident that someone can have the *conception* of Superman without having the conception of Clark Kent. This requires having a MOP that represents Super-Kent but which appears in structured strings of MOPs that express beliefs having such Russellian contents as *Super-Kent came from outer space*

[22] This point is of special importance. A subject attributing beliefs to others can track causal differences correlated with word forms (and correlated with MOPs associated with those word forms) without having any very definite beliefs about the causal mechanisms involved. In other words, my suggestion in no way presupposes that the average speaker has ever considered or explicitly represented facts about the nature of MOPs or the role they play in producing behavior; neither does my suggestion presuppose that any of the subjects involved have introspective access to many facts about their own MOPs, qua nonsemantically individuated units participating in cognitive processing (cf. Rupert, 1998, note 7). Bear in mind also that the very practice of mental-state attribution is arguably driven by the goal of explaining and predicting the behavior of others (and ourselves, by some accounts—see Gopnick (1993)), so the idea that, in attributing beliefs, the speaker has causal-explanatory interests in mind is not at all implausible.

[23] Again, $t1$ and $t2$ are place-holders in a schema that applies to each subject, even though the actual values of $t1$ and $t2$ may be different for different subjects. If, for example, we individuate MOPs neurally, it makes perfect sense to claim that each subject has neural structures that play the role of $t1$ and $t2$ without having to claim that any two subjects share the relevant neural structures.

and *Super-Kent has X-ray vision*. A subject that possesses this kind of MOP, however, need not also possess a MOP that represents Super-Kent and that appears in structured strings expressing such Russellian belief-contents as *Super-Kent is mild-mannered* and *Super-Kent wishes Lois would notice him*. Thus, if 'concept' means 'atomic mental representation', then any subject who has a concept of Superman also has a concept of Clark Kent. But, if we have conceptions in mind, then a subject can possess a conception of Superman without possessing a conception of Clark Kent.[24]

6. The illusion of content-based explanation

Although views of the sort advocated above have some proponents, the approach has not exactly caught fire. One source of resistance has been a sense that differences between MOPs are best characterized *epistemically* (Prinz, 2002, p. 96). When one thinks of Twain-Clemens *as Twain*, one grasps a host of epistemic connections: inferences to Twain-Clemens's attributes as an author and as a pillar of Americana. Call this reaction 'the Semantic Intuition'. These epistemic connections—inferences, beliefs, and the like—seem to be essential to our thinking about Twain-Clemens *as Twain*, and essential to the explanation of behavior in Frege cases. The difference between the belief that Clemens wrote *AHF* and the belief that Twain wrote *AHF* is explicable in terms of content-based and, one might even claim, content-constituting inferences, and furthermore, such differences must ground our explanation of differential behavior in Frege cases.

Contrast this with the views laid out in the two preceding sections. There it was suggested that the difference between two beliefs expressing the same Russellian content can consist entirely in the differential tokening of co-referring vehicles; this approach seems completely disconnected from our intuitive reaction to Frege's puzzle. In Prinz's words, "[I]t [Fodor's approach] fails to cohere with our practice of distinguishing cognitive contents by appeal to semantically interpretable differences in mental states" (Prinz, 2002, p. 97), where 'cognitive content' serves as a place-holder for whatever conceptual content (e.g., a Fregean sense) is distinct from the referential component of content.

A related concern applies in the realm of psychological explanation. We normally advert to content-laden mental states to explain behavior. To explain the subject's behavior, we cite the *reasons* for her behavior; and reasons essentially involve propositions—i.e., the contents of the beliefs, desires, etc. involved (cf. Arjo, 1996, p. 244).

[24] This is not to say that conceptions are well-behaved across contexts or subjects. It may be that our intuitions are right only up to a vague point—one can have *a* Superman conception without having *a* conception of Clark Kent—even though there is no definitive cluster of mental contents counting as *the* conception of Superman or as *the* conception of Clark Kent.

My response comes in three parts. First, I argue that on one plausible account of the linguistic source of the Semantic Intuition, it is illusory. It is true that we draw different inferences from different beliefs and different belief-attributions, even where Russellian content remains constant. This, however, grounds a rejection of the Russellian view only if those differences in inference constitute a difference in belief-content. I argue that the way we learn language creates the illusion that inference partly constitutes content, and this results in the misleading view that where differences in actual or potential inferences do explanatory work, a difference in content is also at work. Second, I argue that concept acquisition, in fact our acquisition of the concept of mental content itself, reinforces this treatment of the Semantic Intuition. Third, I argue that, even in Frege cases, overall psychological explanation makes *some* appeal to content, albeit Russellian content; together with the first two points, this third point will, I hope, appease whatever remaining sympathy the reader feels toward the Semantic Intuition.

6.1. On the linguistic sources of the Semantic Intuition

A number of strategies and methods are at play in language-learning, many of which exploit reliable inferential connections. As a result, when we attribute beliefs differentially in the cases that give rise to Frege's puzzle, we mistake our attempt to track differences in the causal relations supporting such inferences for an interest in content itself.

Much language-learning involves the pairing of word forms with concepts. The child uses a host of pragmatic cues to guide this process (Bloom, 2000). This often involves a fair amount of inference on the child's part, in particular, inferences concerning which properties objects in a given category are likely to possess and, perhaps more importantly, which properties the teacher is likely to have in mind. In this context, the child's acquisition of words is guided by theories, beliefs, or expectations (the precise status is unimportant here) concerning which properties which kinds of things are likely to instantiate and which properties the teacher is likely to think go with which objects. Imagine a parent trying to teach a child a new word—say, 'fep'—and the parent points to a zebra, saying "look at the fep zebra." The child is unlikely to interpret 'fep' as meaning 'striped' because all zebras are striped. So long as the child takes the parent to know that the child knows the meaning of 'zebra' (which may be a reasonable expectation given the child's developmental stage and previous interactions with the parent), the child will assume that the parent would not bother uttering something redundant. Thus, she will look for some salient, but not "definitional," property instantiated by the zebra at hand. In contrast, "Look at that fep animal" may lead the child to associate 'fep' with her visual image of striped-*ness*; in fact, not all animals are striped, and so it is much more likely that the speaker is attempting to draw the child's attention

to the zebra's stripes (Mintz, 2005, pp. 17–18, 22). In a related vein, four-year-olds have been shown to exhibit more robust word-learning when given verbal information about the word, rather than merely being shown a picture (DeBaryshe & Whitehurst, 1986). This reflects a commonsense point about word-learning that applies to adults as well as children. Being shown an example of an echidna may help the uninitiated to understand 'echidna', but adding a verbal description of the echidna's important properties is even more effective. It will help the learner to identify echidnas reliably to have her attention directed to the properties diagnostic of *echidna-hood*, in contrast to what might be idiosyncratic properties of the observed sample.

My point here is that language-learning—learning to use the very words that appear in the attribution of beliefs and desires to others—typically involves learning a host of inferential connections, either of the universally valid or of the merely locally reliable sort. Learning inferential relations is central to the language-learning process by facilitating successful use and application of the terms being learned. It is no surprise, then, that our intuitions about the *contents* of those words are deeply tied to the words' (or their mental representations') epistemic liaisons (Fodor, 1998). It is true that a fair amount of language-learning proceeds by ostension. Nevertheless, consider what is likely to be a central and highly useful rule: "When different sets of inferences are drawn from two different words (e.g., adults say contrasting things when using these words), the words refer to different things." A rule of thumb of this kind is almost certainly at play in language acquisition—if nothing else, as an application of Leibniz's law together with the child's theory of mind.[25]

Prinz (2002, p. 96) claims, "If epistemic differences and similarities play such a central role in verifying our attributions of cognitive content and formal differences do not, then cognitive contents must be epistemic in nature." I think that the preceding discussion undermines this inference. Epistemic liaisons—inferences, theories, beliefs—do, in fact, play an incredibly important role in our acquisition of linguistic items and in our ability to apply them correctly. Nevertheless, this guarantees nothing about the content of the terms acquired, not even that there is such a thing as cognitive content. Referential content is fixed causally-informationally, with inference typically playing a central, mediating role. Inferential connections frequently guide the learning and application of the referring terms that appear in belief attributions; and we typically cash differences between such terms by emphasizing inferential connections. As a result, intuition suggests that the content of an attributed belief differs when terms with different epistemic liaisons are inter-substituted, for we know that the inferences we draw from different words tend to differ. Nevertheless, if we can do with a causal-information theory of content, together with an understanding of the misleading effects of the mediating role of inferential connections, we may have no need for cognitive *content* at all.

I have emphasized the side-effects of strategies used in the teaching and learning of language. In response, the defender of the Semantic Intuition might insist that her intuition is about *beliefs themselves*, not about the words used to attribute them; in which case, the fact that we treat our words a certain way may seem irrelevant. It is time, then, to discuss concept acquisition.

6.2. Concept acquisition and the semantic intuition

Let us begin by considering an important connection between the linguistic and conceptual cases. Concept formation, as well as the acquisition of referring terms, may serve many purposes. Clearly, though, both kinds of representation serve as the basis for induction and related forms of inference. One learns that bats are mammals; given, then, one's background knowledge concerning mammals, if one hears the term 'bat' applied to an unfamiliar-looking animal, one will infer that this new critter has a heart (cf. Markman, 1989, pp. 95–101).[26] Thus, it seems plausible either *(1)* that developing subjects assume that what appears true of linguistic meaning—its connection to inference, for example—is also true of conceptual content or *(2)* that young (and perhaps not-so-young) subjects develop inference-related intuitions about linguistic meaning *because* they have an existing, though perhaps implicit, inference-based view of conceptual content. The latter might result from whatever limited introspective awareness subjects have of their own cognitive processes in which inferences are drawn on the basis of category membership (cf. Margolis & Laurence, 2003). In support of *(2)*, we might wonder how often a well-designed cognitive system would have occasion to token a single, simple representation, without inferring its application from the application of some other term or doing the converse. If either case *(1)* or *(2)* obtains, it is likely that factors similar to those

[25] Of course, kids can get carried away here. Consider the parent who must insist to the child "Uncle Henry *is* my brother," because the child has learned the typical properties of a brother in a different context, where different properties are salient, from that in which she has learned the typical properties of an uncle. As a result the child mistakenly thinks 'brother' and 'uncle' have mutually exclusive extensions (i.e., refer to different collections of things) and must be talked out of this misapprehension. On a more philosophical note, consider the possibility that conscious experiences *just are* brain states. If this is true, we are misled about the identity by our use two different clusters of terms to talk about these states, cluster of terms that enter into rich, nonoverlapping inferential networks. If content is fixed by causal relations to the world, and if the terms involved (and their corresponding mental representations) come into the right causal relation to their referents via mediation by distinctive inferential networks, it will seem as if the terms in question—say, 'conscious experience of redness' and 'activation of such-and-such cells in V4'—*must* refer to different things (cf. the various objections to mind-brain identity theory considered by Smart (1959)).

[26] Note in this connection the extent to which successful inference is treated, both by parents and researchers (e.g., Liu, Golinkoff, & Sak, 2001, p. 1686), as a criterion of term and concept acquisition.

identified in connection with linguistic meaning drive the Semantic Intuition as it applies to mental content.

Turn now to issues related specifically to concept acquisition. As noted earlier, even if a causal-informational theory of mental content is true—as I think it is—the process of concept acquisition can be inferentially mediated, where this includes mediation by socially guided inference; insofar as the process involves the exploitation of inferential connections and insofar as the presence of such processing shapes our *conception* of content (or the production of intuitions about content), we come to think of concept acquisition as itself the acquisition of clusters of inferences.[27] The role of social mediation is likely to be misunderstood, so some caveats are in order. First, I do not mean to deny the innateness of many mental representations. Second, I do not mean to assert that the developing child "magically" soaks up her culture through language. (And note the connection between these two issues: no one without any mental representations can acquire concepts through linguistic interaction; for, absent representations of the incoming speech stream and the environment to which that speech is keyed, the culture surrounding a child remains static and noise.)

Because the innateness of concepts seems to preclude a social role in their acquisition, the question of nativism merits further discussion. I take it that, generally speaking, the acquisition of even robustly innate concepts involves social interaction, at least for the purpose of triggering those concepts (Fodor, 1981). This applies all the more fully to concepts that are only weakly innate. Let me explain. Fodor (1975, 1981) has famously argued that a surprising number of concepts are innate, but at least in some forms, his argument rests on a particularly weak standard of innateness—that of not being learned. If something's being non-learned suffices for its being innate, and if learning essentially involves the formulation and testing of definitional hypotheses, then all of our concepts that cannot be satisfactorily defined are innate; for if they cannot be defined, then no hypothesis can be formulated that captures their meaning. Fodor concludes that humans possess a surprisingly large set of innate concepts, a set including such concepts as CARBURETTOR and JUSTICE. Many of these concepts, however, fail other compelling tests for innateness. It is not the case that one is born with these concepts up and running. It is not the case that one would develop these concepts across a wide range of environments and stimuli. Furthermore, "weakly" innate concepts that are, in fact, acquired by a given subject may be few among a vast array of concepts the acquisition of which is consistent with that subject's genome but the lion's share of which the subject does not acquire. Thus, to explain how these weakly innate concepts are acquired, we should appeal to contingent facts about the subject's environment and the subject's interaction with that environment. Such details partly account for the child's coming into causal contact with the world in such a way that she acquires only a small portion of the non-definable concepts from the enormous range of nondefinable concepts she could possibly acquire. I would maintain that a large part of this story involves mediation by inferential relations among concepts (Rupert, 2001).

Think of this a bit differently. The matter and structure of the human brain may render it unlikely that *any* portion of the brain will become causally sensitive to certain properties while rendering it more likely that some portion or other of the brain will become causally sensitive to other properties. Of those properties in the latter category, concepts of them that do not pass standard tests for innateness are weakly innate (if innate at all). Now add a causal-informational semantic theory and the assumption that having the content appropriate to concept C is necessary to acquiring C. Given what we know about the brain, in order for a neural structure or class of neural structures to become a reliable detector of property P, that structure has to interact with instances of P (and this holds regardless of whether the brain is predisposed to have parts that reliably detect P). This epigenetic interaction constitutes part of the process by which passing noise and cognitively insignificant patterns of activation come to be the neural relata of *stable* content-relations. This process can be both socially and inferentially mediated *and* result in the acquisition of nondefinable, atomic concepts.[28]

In the present context, it is important to bear in mind the contribution of social factors to the causal processes that "recruit" neural units and stabilize their relation to environmental properties and individuals these vehicles come to represent—social factors that can operate even

[27] Humans are unaware of many of inferential processes taking place in their own cognitive systems (see, e.g., Nisbett & Wilson, 1977). My suggestion in the text, however, does not rely on a robust thesis of introspective access to cognitive processing. Rather, I can rely on the more limited claim that subjects experience some significant range of the inferential processes involved—enough to mislead them into thinking that inferential relations are indicative of content. Notice that some of the inferences at issue are explicitly drawn in verbalized language by a teacher; the learner's intake of this does not provide a very compelling example of cognitive processing to which subjects have little or no general-purpose or conscious access. The issue of introspective access may, though, be a red herring. My proposal does not require that the subject be consciously aware of, or that the subject be able accurately to report on, *any* of the inferential processes involved. Assume, plausibly enough, that relatively inaccessible mechanisms generate the Semantic Intuition. Those mechanisms can be "trained up"—'fooled' might be better—by inferential processes occurring in language and concept acquisition to which the subject has little access. In this way, the entire process described in the text to account for the Semantic Intuition might operate at a relatively inaccessible level.

[28] For more on the way in which environmental events causally interact with the flux of events in the brain so as to stabilize vehicles that can stand in the content-fixing causal relations, see Rupert (1998, 1999, 2001). Dan Weiskopf (forthcoming) has recently emphasized the possibility that *purposive* activity can help to recruit neural structures so that they become cognitive vehicles.

in the case of innate concepts. Many mental representations refer to properties or kinds the instances of which vary significantly along perceptual dimensions. Being given labels for such properties allows the learner to develop stable detectors of them. The provision of such labels causes the learner to form mental representations of the words themselves, which can then act as anchors for internal processing (Clark, 1998; Gentner, 2003). The activation of the mental representation of a word can cue the activation of an emerging vehicle—distinct from the mental representation of the word itself—that eventually will carry the content to be learned. Simultaneously, the mental representation of the word activates other representations, e.g., of properties observed in the past in connection with that word form. Some of these properties are specifically pointed out by parents; Gentner (2003, p. 200) gives the example "That's a wolf. It's like a dog, except it's wilder. As Gentner argues, much concept acquisition proceeds by internal comparison (which is, of course, not definition). Such comparative processes require at least that some mental representations already be in place, but language frequently directs the learner's attention to the relation between the existing mental representations the comparison of which causes the acquisition of a new concept. For my purposes, "directing attention to" need amount to nothing more than causing the subject to activate two or more representations the co-activation of which is likely to cause the coining of a MOP, or to cause the further stabilization of one that has previously been coined; furthermore, coining involves nothing more than the strengthening or weakening of neural connections such that the neural unit in question becomes more likely to participate in cognitively relevant causal processes. Many concepts—of relations between relations, for example—are simply too difficult to acquire without the anchoring representations provided by language, regardless of what the nativist status of these concepts might be.

Furthermore, the very availability of some mediating structures may be a product of many small steps spread out over generations and passed on by culture: a crude (possibly linguistic) artifact can causally mediate the acquisition of a concept that would not have been acquired otherwise. This concept can contribute to a process of reasoning that eventuates in the construction of a more sophisticated artifact, which can then causally mediate the acquisition of a concept that would not have been acquired if only the crude artifact had been available; and so on.

I submit that the processes I have described produce a mistaken intuition: that when confronted with differences in inferential roles (which are sometimes internally represented patterns of inference in speech), these differences result from differences in the content of the component mental representations. The cognitive system possesses various monitoring systems, some, but certainly not all, of which are associated with consciousness or generalized cognitive access (see note 27, above). To the extent that these systems record what is going on in the process of concept acquisition, subjects will be given the illusion that acquiring a concept—"getting the idea"—involves the cementing of certain inferential relations. If a neural unit's becoming a reliable detector involves causally interacting with examples of what is to be detected, and if that interaction involves mediating structures, it is no wonder that the cognitive system acquires the disposition to treat the mediating structures—including inferential relations—as somehow central to the concepts so acquired.

To strengthen this case, consider the acquisition of the very concept of mental content. On the sort of causal-informational view I have endorsed, mental content is fixed by causal relations, typically to organismically external properties, kinds, and individuals. What intuitions will the mind itself develop about this content-fixing process? The cognitive system engages in various forms of self-monitoring, including the monitoring of the self's interaction with the environment. There is, however, little reason to think that these forms of monitoring will divulge (that is, lead to the construction of true beliefs about) the content-fixing process itself. By nearly all accounts, the nature of the causal-informational relation that fixes the content of mental representations is somewhat complex and difficult consciously to home in on. It may be much easier to home in on the content-fixing causal relation via an indicator of its obtaining than by the subject's defining it in terms of concepts she already possesses. Notice this former kind of causal resonance does not require direct access to the external world. Rather, the subject coins a MOP that is or becomes differentially causally responsive to her internal units that refer to objects or properties in her immediate environment—in contrast to those that do not. Perhaps this is learned quite early, in the distinction between mental representations of fictional things and mental representation of real things. In this way, the child might develop an atomic conceptual unit tracking the content-fixing relation without having any *conception* of that relation and without needing any direct access to the external relata or the causal relation itself.[29] That is, she can coin a MOP, call it CONTENT, that tracks the causal-informational content-fixing relation by its being tokened differentially in response only to the local relata (those internal to the cognitive system) that stand in the causal-informational content-fixing relation. I have put this in terms of the child's acquisition of a notion of content, but if that seems implausible to the reader, it might be thought that the concept of content is acquired only later, in, say, philosophy or psychology class (cf. Margolis & Laurence, 2003). Either

[29] Up to this point, I have used names for fictional characters as if their fiction-related status were unimportant: I treated the Superman fiction as real. Now we encounter a specially important role for the distinction between fictional referents and those referents to which the thinker is connected by causal chains consisting of concrete events. Nevertheless, the essential points of the discussion of Super-Kent can easily be translated into the language of a non-fictional example, that of Twain-Clemens, for instance.

way, though, my point is this: it is plausible that at some point the subject could coin a MOP that bears the right causal relation to *the content-fixing causal relation itself*, without the subject's needing to have direct access to (in the sense of having a MOP that *directly* causally interacts with) the external relata standing in that relation to the subject's referring MOPs or to the causal interaction that fixes content.

The preceding story does not make it clear how the concept of content comes to be associated with the concept of inferential relations. Consider, then, that associative mechanisms influence the functioning of the cognitive system, as do certain hypothesis-formation mechanisms (though neither of these need involve conscious access), and this can cause the cognitive system to develop robust associations between CONTENT and the system's MOPs representing certain kinds of patterns of inference. Approach this from the developmental perspective. Let's say the child begins to acquire CONTENT by noticing the following difference. One gets connected to real objects by encountering their properties in sensory observation or by having their properties described; in either case, it is clear that the properties involved provide *epistemic access to the things being referred to*. Identifying these properties grounds our successful application of the terms being learned (Fodor, 1998). In contrast, in the acquisition of fictional names or concepts of fictional individuals, kinds, or properties, the purported referents are encountered by the child (or described to the child) in a way that does not presuppose epistemic access; the properties associated with the name or mental representation are not offered as ways to identify correctly the individual in question or instances of the property or kind in question. Identification of the actual referents is out of the question; no such things exist. Granted, in order to follow a story, the child has to be able to recognize which character the parent is talking about at a given time—the child must be able quickly to infer "x is winged" from "Pegasus $= x$"—but the inferences involved cannot guide the child's own interaction with those characters; they cannot facilitate the success of the child's plans and projects as she "interacts" with those characters in the way inferences do when they facilitate the child's successful interaction with actual objects and their properties.

As a result the child (or adult) acquires the two related "beliefs" with roughly the following content. First, concepts that actually refer to individuals, properties, and kinds found in the environment have a special status (their activating CONTENT partly constitutes the child's recognition of this status). Second, concepts with content have their use mediated by the actual application of other concepts to things in the environment. The result is the association-based illusion (or mistaken hypothesis) that content is somehow specially related to the process of mediated application, which is a form of inference. The parent asserts, "That's a zebra," where 'zebra' is, so far as the child can tell, being applied to a concrete thing in the environment. The parent then says, "Zebras have stripes," which gives the child epistemic access to zebras via one of their properties; this results in the child's use of one mental representation, STRIPED, to govern the tokening of another, ZEBRA. Thus, in the case of atomic MOPs that have Russellian content, inferences matter much more than they do in the case of atomic vehicles that do not. Inferences allow the subject access to actual objects—to identify them in actual world, as opposed to being able to follow the fiction—thus creating a sense that the inferences are related to contentful concepts in a way that they are not in the cases of non-contentful concepts. In the former case, the world thrusts upon us a reliance on inferential mechanisms; in the latter, the inferences have something more like the character of stipulation. Furthermore, it is, arguably, the very application of inference in the former manner that helps the child to form the distinct categories of fictional terms, on the one hand, and genuinely referring terms, on the other, and the marking of this distinction helps to cause CONTENT to come into the content-fixing relation to the content-fixing relation itself.[30]

When the cognitive system acquires a concept of mental content, it represents the actual causal relation involved. It does so, however, via an atomic vehicle's coming into the correct causal relation to the *content-fixing causal relation itself*. This alone gives the cognitive system no particular ability to elaborate on that content—that is, to develop detailed, true beliefs in which the concept of content appears as in the subject position. Rather, the cognitive system is more likely to be misled on this score. The acquisition of the concept of content is causally mediated

[30] There is an affinity between some of what I have said and what Fodor (1998) and Margolis and Laurence (2003) say about intuitions of analyticity. Notice, however, the important distinction between the rather vague, meaning-related association I have in mind and intuitions concerning which specific inferences are to be counted as analytic. I am interested in our inclination to associate (not necessarily analytically) the concept of inference with the concept of content. So, although Fodor feels pressed to explain why some epistemic liaisons seem meaning-constitutive and others do not—appealing specifically to examples where the application of a concept is governed by only one other criterion—I mean to explain only why someone would mistake for a semantic matter the indirect causal influence of syntactic vehicles on inference. Thus, my account need not treat single-criterion concepts any differently from the way in which it treats multiple-criterion concepts; nor need I claim that there is any tractable way to make sense of that distinction. If there is any category of special cases, on my view, they are the most immediate sensory concepts. These are, as we might say, *zero*-criterion concepts. It is plausible that, at least beyond the inner workings of the sensory modules, one does not infer the application of sensory representations from any other mental representation; one simply *looks* (feels, etc.). So, my speculative story about the association between content and inference does not apply to these cases, but neither are they cases in which it is at all easy to construct Frege cases—at least not where *both* MOPs are immediate sensory representations. Now, we might think of, e.g., red in two different ways—Jackson's Mary is alleged to do so (Jackson, 1982)—but it seems implausible that someone would think of a single shade of red under *two* different immediate sensory concepts, i.e., where both MOPs are atomic, immediate sensory concepts. It is certainly not this sort of case that has given rise to Frege's puzzle or to concerns about Frege cases (for further discussion of this last point, see note 32, below).

by various inferential processes that, in particular cases, mediate the acquisition of garden-variety concepts; these are the processes available to the cognitive system (consciously or not) when stabilizing a concept of content. Here the association between inference and content is cemented by the privileged role inferences play in marking referring terms, which marking allows the acquisition of the concept of content. As a result, our mental representation of content is more closely tied—by subjective association, anyway—to our concept of inferential relations than it is to our concept of one thing's carrying information about another or one thing's being reliably correlated with another (and thus it is a lot of work to discover the actual nature of the content-fixing relation). As a result, when we see differences in inferential relations, we assume they reflect differences in content; in other words, we develop the compelling, but misguided, Semantic Intuition.[31]

6.3. Content-based psychological explanation

In the case of psychological explanation, something of the Semantic Intuition can be salvaged; for in virtually all Frege cases, the correct explanation of behavior appeals to some rationalizing content; this follows from the quasi-syntactic view. I elaborated on the picture presented by Fodor and Schneider in the way I did partly to elucidate the role content plays in psychological explanation even in Frege cases, including interpersonal Frege cases. In the example involving Super-Kent and the attacking villain, the difference in behavior among the two groups of bystanders is explained partly by the fact that members of each group exhibit among themselves a shared contrast between pairs of terms (but not shared terms themselves). To explain fully the actual behavior that ensues one must appeal to the *contents* of the mental representations to which one of those contrasting MOPs is causally connected in the cognitive system, i.e., causally connected via the MOP's causal connection to the realizers of the other mental representations in question. The activation of one MOP rather than the other might, for example, be caused by observation of the property of being bespectacled. This causes running (or a failure to ask for help) because it causes the activation of such MOPs as MILD-MANNERED and WIMPY, which themselves appear in LOT strings that express beliefs such as "wimpy people aren't much good against super-villains". This belief appears as a premise in a practical syllogism the conclusion of which is "run away now". This preserves an important role for content-based explanation, even in Frege cases.[32]

7. Conclusion

I have claimed that in typical cases, belief is a relation to the worldly objects and properties referred to by the terms used to attribute such beliefs. The propositions believed are Russellian. Nevertheless, the *entire* story of belief-attribution and belief-related effects on behavior adverts to mental representations conceived of nonsemantically—i.e., to the cognitive vehicles of Russellian content.

I have spent much of my time addressing a powerful objection to this view, an objection rooted in the Semantic Intuition. In response, I have appealed to causal-informational theories of content, which allow substantial mediation in the fixation of mental content but do not treat the mediating mechanisms as constitutive of the content so fixed. This point was complemented by an account of how such mediation can play a role in language-learning and can influence our understanding of linguistic and conceptual content. Such influence creates an illusory intuition, one holding that our belief-attributing tendencies and our use of beliefs to explain behavior *must* be understood in semantic terms. At points, I have emphasized the social aspect of such mediation, which I take to be practically indispensable if humans are to acquire a rich language and a wide range of concepts; this social contribution—coming largely in the form of the social guidance of inferences drawn during language and concept acquisition—helps to produce the misleading Semantic Intuition.

I close with some comments concerning what has and has not been accomplished. The view developed herein offers cognitive scientists and philosophers of cognitive science much of what they should want from a theory of content

[31] Consider a complication: Mary will insist that, before someone she trusted told her "Mark Twain is Samuel Clemens," she did not believe it. Some readers might find it amazing, and amazingly implausible, that Mary could be wrong about this. Nevertheless, differences in the patterns of her inferences involving her own mental representations can cause her falsely to deny that she held the belief in question. Previously, the two relevant MOPs entered into distinct causal-inferential relations; she mistakenly thinks that such differences amount to differences in content; thus she concludes that she had two distinct beliefs with conflicting contents about Twain and Clemens. She might even think that attributing beliefs with distinct content best explains her earlier behavior, given that she thinks, wrongly, that a belief in the true identity claim would have led to different behavior than she actually exhibited.

[32] Are there Frege cases where content-based explanation plays *no* role, e.g., cases involving atomic perceptual representations whose effects on behavior are unmediated? Perhaps, although I'm a bit skeptical; it would seem that all forms of intelligent behavior are mediated to some extent by background beliefs, knowledge, etc. In contrast, reactions involving no mediation seem to be mere reflexes, with regard to which our intuitions about rationalizing explanation do not gain much purchase. There could, however, be complicated cases involving two mutually disconnected structures of interrelated beliefs, all of whose cross-structure, corresponding components are "Fregefied." Absent a realistic example, this possibility seems to me beyond the purview of the intuitive objection at issue (i.e., I feel little pressure to respect the Semantic Intuition with regard to it or to respect the related claim that human behavior must be given a rationalizing explanation). Given a realistic example, the Russellian psychologist will, I suspect, treat it as a case where the Semantic Intuition is erroneous. Just as in the cases discussed earlier, the application of the concepts involved would involve mediation, providing the illusory sense that a difference in content separates the MOPs in question; nevertheless, the Russellian psychologist might have to treat it as a case where only a syntactic explanation of behavior can be given.

as well as from a solution to Frege's puzzle. It is naturalistic, grounded in what appear to be fundamental physical relations such as causality and the carrying of information. It also respects the individualistic methodology that has been so successful in cognitive science, while simultaneously accommodating the highly social nature of language-learning and concept acquisition. I cannot, however, pronounce in conclusion, "Let Russellianism reign!" To do so would be to slight the numerous further objections that have been raised to the Russellian view, many of which arise from work on the semantics of belief-attributions (McKay & Nelson, 2005). I certainly cannot canvass and respond to all of these in the space allotted. Nevertheless, I hope that my discussion has been a source of some optimism for readers inclined toward a broadly Russellian view.

Acknowledgements

I would like to thank Graeme Forbes, David Barnett, Bill Ramsey, Clayton Lewis, and the referees for *Cognitive Systems Research* for their insightful comments on an earlier version of this essay. Thanks also to Murat Aydede and Jerry Fodor for helpful e-mail exchanges, to Susan Schneider and Chris Heathwood for useful discussion of some of these issues, and to Leslie Marsh for editing this special issue of *CSR*. I wrote the first draft of some of this material while I held a National Endowment for the Humanities Fellowship for College Teachers. My thanks to the NEH for their support.

References

Arjo, D. (1996). Sticking up for Oedipus: Fodor on intentional generalizations and broad content. *Mind & Language, 11*, 231–245.

Aydede, M. (1998). Fodor on concepts and Frege puzzles. *Pacific Philosophical Quarterly, 79*, 289–294.

Aydede, M. (2000). On the type/token relation of mental representations. *Facta Philosophica: International Journal of Contemporary Philosophy, 2*, 23–49.

Block, N. (1986). Advertisement for a semantics for psychology. In P. French, T. Uehling, & H. Wettstein (Eds.). *Midwest studies in philosophy. Studies in the philosophy of mind* (Vol. 10, pp. 615–678). Minneapolis: University of Minnesota Press.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge: MIT Press.

Braun, D. (2001). Russellianism and prediction. *Philosophical Studies, 105*, 59–105.

Burge, T. (1979). Individualism and the mental. In P. French, T. Uehling, & H. Wettstein (Eds.), *Midwest studies in philosophy IV* (pp. 73–121). Minneapolis: University of Minnesota Press.

Chalmers, D. (1994). On implementing a computation. *Minds and Machines, 4*, 391–402.

Clark, A. (1998). Magic words: How language augments human computation. In P. Carruthers & J. Boucher (Eds.), *Language and thought: Interdisciplinary themes* (pp. 162–183). Cambridge: Cambridge University Press.

Cummins, R. (1996). *Representations, targets, and attitudes*. Cambridge: MIT Press.

DeBaryshe, B. D., & Whitehurst, G. J. (1986). Intraverbal acquisition of semantic concepts by preschoolers. *Journal of Experimental Child Psychology, 42*, 169–186.

Devitt, M., & Sterelny, K. (1987). *Language and reality: An introduction to the philosophy of language*. Cambridge: MIT Press.

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

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge: MIT Press.

Fodor, J. A. (1975). *The language of thought*. Cambridge: Harvard University Press.

Fodor, J. A. (1981). The present status of the innateness controversy. In J. A. Fodor (Ed.), *Representations* (pp. 257–316). Cambridge: MIT Press.

Fodor, J. A. (1987). *Psychosemantics: the problem of meaning in the philosophy of mind*. Cambridge: MIT Press.

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

Fodor, J. A. (1990b). Substitution arguments and the individuation of beliefs. In Fodor, 1990a, pp. 161–176.

Fodor, J. A. (1994). *The elm and the expert: Mentalese and its semantics*. Cambridge: MIT Press.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford: Oxford University Press.

Forbes, G. (1989). Cognitive architecture and the semantics of belief. In P. French, T. Uehling, & H. Wettstein (Eds.), *Midwest studies in philosophy XIV* (pp. 84–100). Notre Dame: Notre Dame University Press.

Frege, G. (1892). Über sinn und bedeutung. In *Zeitschrift für Philosophie und philosophische Kritik* (Vol. 100, pp. 25–50). Translated as On Sense and Reference by M. Black in Translations from the Philosophical Writings of Gottlob Frege, P. Geach and M. Black (Eds. and Trans.) (3rd ed.,) Oxford: Blackwell, 1980.

Gentner, D. (2003). Why we're so smart. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 195–235). Cambridge: MIT Press.

Gibbs, R. (2006). *Embodiment and cognitive science*. Cambridge: Cambridge University Press.

Gopnick, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences, 16*, 1–14.

Grice, H. P. (1957). Meaning. *Philosophical Review, 66*, 377–388.

Grush, R. (1997). The architecture of representation. *Philosophical Psychology, 10*, 5–23.

Hall, N. (2000). Causation and the price of transitivity. *Journal of Philosophy, 97*, 198–222.

Harnad, S. (1990). The symbol grounding problem. *Physica D, 42*, 335–346.

Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly, 32*, 127–136.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge: Harvard University Press.

Lewis, D. (1973). Causation. *Journal of Philosophy, 70*, 556–567.

Liu, J., Golinkoff, R. M., & Sak, K. (2001). One cow does not an animal make: Young children can extend novel words at the superordinate level. *Child Development, 72*, 1674–1694.

Margolis, E. (1998). How to acquire a concept. *Mind & Language, 13*, 347–369.

Margolis, E., & Laurence, S. (2003). Should we trust our intuitions? Deflationary accounts of the analytic data. *Proceedings of the Aristotelian Society, 103*, 299–323.

Markman, E. (1989). *Categorization and naming in children*. Cambridge: MIT Press.

McDermott, M. (1995). Redundant causation. *British Journal for the Philosophy of Science, 46*, 523–544.

McKay, T., & Nelson, M. (2005). Propositional attitude reports.*Stanford Encyclopedia of Philosophy*, Edward N. Zalta (Ed.), <http://plato.stanford.edu/entries/prop-attitude-reports/>.

Millikan, R. G. (1984). *Language, thought and, other biological categories: New foundations for realism*. Cambridge: MIT Press.

Mintz, T. H. (2005). Linguistic and conceptual influences on adjective acquisition in 24- and 36-month-olds. *Developmental Psychology, 41*, 17–29.

Nisbett, R., & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259.

Port, R., & van Gelder, T. (Eds.). (1995). *Mind as motion*. Cambridge: MIT Press.

Prinz, J. J. (2002). *Furnishing the mind: Concepts and their perceptual basis*. Cambridge: MIT Press.

Putnam, H. (1975). The meaning of meaning. *Mind, Language, and Reality: Philosophical Papers* (Vol. 2, pp. 215–271). Cambridge: Cambridge University Press.

Rupert, R. D. (1996). The best test theory of extension. Ph.D. dissertation, University of Illinois at Chicago.

Rupert, R. D. (1998). On the relationship between naturalistic semantics and individuation criteria for terms in a language of thought. *Synthese, 117*, 95–131.

Rupert, R. D. (1999). The best test theory of extension: First principle(s). *Mind & Language, 14*, 321–355.

Rupert, R. D. (2001). Coining terms in the language of thought: Innateness, emergence, and the lot of Cummins's argument against the causal theory of mental content. *Journal of Philosophy, 98*, 499–530.

Ryder, D. (2004). SINBAD neurosemantics: A theory of mental representation. *Mind & Language, 19*, 211–240.

Salmon, N. (1986). *Frege's puzzle*. Cambridge: MIT Press.

Schneider, S. (2005). Direct reference, psychological explanation, and Frege cases. *Mind & Language, 20*, 423–447.

Smart, J. J. C. (1959). Sensations and brain processes. *Philosophical Review, 68*, 141–156.

Stampe, D. W. (1977). Toward a causal theory of linguistic representation. In P. French, T. Uehling, Jr., & H. Wettstein (Eds.), *Contemporary perspectives in the philosophy of language* (pp. 81–102). Minneapolis: University of Minnesota Press, 1979.

Weiskopf, D. On the origin of concepts. *Philosophical Studies*, forthcoming.