Psychosemantics and the Rich/Thin Debate[*]

E. J. Green

(forthcoming, *Philosophical Perspectives*)

Abstract: I consider the issue of whether perceptual content is "rich" or "thin," using the lens of psychosemantics. As a case study, I examine Neander's (2017) recent psychosemantic theory of perceptual representations, which supports a thin view of perceptual content. I argue that the view faces difficulties, and that these difficulties trace directly to the component that makes it thin-friendly. I show that this sort of issue is not unique to Neander's theory—it also arises for Dretske's and Fodor's accounts. I then articulate a more general challenge for any psychosemantic theorist seeking to retain a systematically thin view of perceptual content. I conclude that a viable psychosemantics of perception is unlikely to support the thin view.

1. Introduction

It's widely agreed that our perceptual systems construct representations of objects and properties in the environment. But which properties do our perceptual systems represent? There is considerable disagreement about this issue. Nowhere is the discord more apparent than in recent disputes over whether perception represents "high-level" properties. The present paper considers this issue through the lens of psychosemantics. I'll argue that any psychosemantic theory that delivers scientifically respectable verdicts about the perceptual representation of low-level properties is likely to also entail that there are perceptual representations of high-level properties.

To fix ideas, consider a recent exchange between Ned Block and Tyler Burge. Block (2014) contends that the visual system constructs representations of faces, facial expressions, and face gender. His argument draws primarily on adaptation aftereffects. There is evidence that extended exposure to an angry face makes a subsequent neutral face look more fearful (an aftereffect) (Butler et al. 2008), and that extended exposure to a "stretched" face makes a normal face look shorter (Susilo et al. 2010). In reply, Burge (2014) argues that this sort of evidence fails to establish that perceptual states represent *facial* properties. The adaptation patterns, says Burge, would be just the same if these perceptual states instead represented low-level configurations diagnostic of facial

[*] Thanks to Alex Byrne, Carla Merino-Rajme, Adam Pautz, Susanna Siegel, and Steve Yablo for helpful comments on earlier versions of this paper.

properties—say, a generic geometrical configuration characteristic of angry faces.[1] Briscoe (2015: 188) articulates a similar line of thought: "[T]he available scientific evidence underdetermines the question of whether the aftereffects observed in face adaptation studies reflect changes in the sensitivity of neurons that code for the lower-level 'building blocks' of possible faces. Hence, it is unclear at present how best to characterize the nature of face adaptation aftereffects."

So we want to know whether vision represents facial properties *per se*, or merely the lower-level diagnostic features of facial properties. This is a difficult issue.[2] To appreciate its difficulty, suppose that we learned *all* the relevant physiological, functional, and formal-syntactic facts about visual processing. Presumably, we would find that a complex series of physiological processes produces a state-type R in response to angry facial expressions. We would also learn some interesting facts about R, such as that it is susceptible to threshold shifts resulting in adaptation, and that it feeds into postperceptual systems that control adaptive behavior in the presence of angry faces. But how do we determine whether R represents *angry face*, rather than a low-level feature configuration diagnostic of angry faces?

One way to try to resolve such disputes is by appeal to psychosemantic principles—principles governing what makes it the case that a mental representation has the content it does. Armed with independently plausible guidelines for determining the contents of perceptual representations, we could use them to figure out what R represents. This procedure could either vindicate the high-level content attribution or undermine it. Psychosemantics, of course, is not the *only* way to investigate issues about high-level content. However, given that it is at least *one* key

[1] Note that Burge is not globally skeptical of the idea that visual states have high-level content. I use this example only to underscore the difficulty in resolving issues about whether specific visual states have high-level content.

[2] Comparably difficult issues come up in debates about the perceptual representation of natural kinds (Siegel 2010; Prinz 2013), causation (Scholl & Tremoulet 2000; Siegel 2009; Rips 2011; Siegel & Byrne 2016), numerosity (Fish 2013; Briscoe 2015), basic-level categories (Fodor 1983; Mandelbaum 2017; Block ms.), and "gist" properties like being a beach or a forest scene (Fish 2013; Bayne 2016; Mandelbaum 2017; Block ms.).

source of evidence concerning the contents of mental states, we should investigate what a plausible psychosemantics is likely to say about the perceptual representation of high-level properties.

As a case study, I'll focus on Karen Neander's (2013, 2017) recent, carefully worked-out psychosemantics of perceptual representation. Neander proposes that the content of a perceptual representation is, roughly, the property or event-type that the representation's producing mechanism functions to respond to. Critically, this view is one of only a handful of psychosemantic theories whose verdicts fall clearly in line with thin views of perceptual content, and it is certainly the most refined. Neander's exposition is also noteworthy for containing substantive arguments *in favor* of limiting perceptual representations to thin contents involving the "surface appearances of things" (Neander 2017: 120; see the Appendix for discussion of these arguments). This makes the view a natural focal point. I'll argue that it faces difficulties, and then I'll argue that these difficulties are representative of a more general challenge facing any psychosemantic theory that seeks to restrict perception to thin content.

The plan is as follows. Sections 2-3 introduce Neander's view and explain why it entails a thin view. Section 4 argues that the element of Neander's view that precludes high-level content also generates implausible verdicts concerning low-level content. Section 5 shows that these difficulties generalize to other thin-friendly psychosemantic theories—namely, Fodor's and Dretske's accounts. Section 6 discusses two potential remedies for the problem, but shows that both require admitting high-level contents into perception. Section 7 draws out a more general challenge for any thin-friendly psychosemantics of perception: The sorts of achievements needed for perceptually recovering high-level properties are not different from those already involved in recovering low-level properties. This makes it difficult for a psychosemantics of perception to rule *out* representations of causation, facial expressions, animacy, and so on, while ruling *in* representations of low-level properties like motion and shape.

Before moving on, I should make a couple of clarifications. First, two debates about high-level content should be distinguished. On the one hand, there is a debate about the representational limits of visual *experience*. This concerns whether visual experiences ever represent high-level properties.[3] On the other hand, there is a debate about the representational limits of the visual *system*. This concerns whether the visual system ever computes over or outputs representations of high-level properties. I'll be focusing on the latter issue.[4] I'll understand the *visual system* to consist just of those visual processes that operate relatively autonomously, so with minimal cognitive influence. Thus, when I speak of thin views in what follows, I'll have in mind views on which these quasi-autonomous portions of visual processing are limited to representing low-level properties. And when I speak of rich views, I'll have in mind views on which this is not true.

Second, we should be careful to differentiate the rich/thin debate from debates about cognitive penetration. Cognitive penetration would of course be one *mechanism* for yielding rich content. Roughly, the idea would be that a perceiver who acquires the ability to cognitively represent a certain high-level category might make that representation available to perceptual processing as well, thereby enriching the representational resources of perception. But there is no reason to think that this is the only—or even the most likely—way that perceptual systems might come to represent high-level content. At least two other routes are available. First, perceptual systems might be hardwired to recover certain high-level properties, such as causation or animacy (Scholl & Tremoulet 2000; Scholl & Gao 2013). Second, perceptual systems might acquire the ability to represent high-level properties via perceptual learning. Below, I'll discuss evidence concerning the perception of

[3] Following Siegel (2010: 7), I'll understand high-level properties to include any properties of an object beyond its shape, size, color, illumination, texture, and motion.

[4] However, the two issues are closely related. Some, such as Byrne (2009), motivate a thin view of visual experience on the basis of a thin view of the visual system's outputs. If the visual system outputs representations of high-level properties, then this way of arguing for a thin view of experience will not work. For present purposes, I'll leave open the question of how (and whether) the contents of visual experiences are grounded in the contents of representations produced by the visual system. One might suggest, for instance, that the contents of representations produced by the visual system outstrip the contents of visual experience. If so, then even if a rich view were true of the visual system, it could be false of visual experience.

causation. I think that *if* causation is indeed perceived, then it is typically perceived by way of hardwired perceptual mechanisms.

2. Causal-Informational Teleosemantics

Traditional psychosemantic theories divide into two categories. First, there are theories on which the facts about a representation's content are grounded in facts about the biological functions of systems that produce or consume the representation (Millikan 1984; Papineau 1998; Price 2001).[5] Second, there are theories on which the facts about a representation's content are grounded in facts about the representation's causal, counterfactual, or informational relations to the environment (Dretske 1981; Fodor 1990; Tye 1995). Neander's (2017) view is notable for combining elements of both of these traditions. Roughly, she holds that a representation R picks out the external condition that R's producing system functions to respond to.

A central notion of the view is that of a *response function*, which Neander characterizes as a function "to respond to something by doing something" (126). More specifically, a system has a response function when it has the function of changing its state in certain ways in response to certain external state changes. For example, the pineal gland has the function of releasing melatonin in response to the dimming of light. Neander's main idea is that our perceptual systems have response functions, and in particular that they have functions to produce representations in response to certain states of the world. She captures this idea with the following principle, labeled "CT" for "<u>c</u>ausal-informational <u>t</u>eleosemantics":

> CT: A sensory-perceptual representation, R, which is an (R-type) event in a sensory-perceptual system (S), has the content *there's C* if and only if S has the function to produce R-type events in response to C-type events (in virtue of their C-ness). (Neander 2017: 151)

[5] The proper function of an item X is, roughly, what items of its type were selected to do. It is what ancestors of X did that led to their survival and reproduction, and thus X's presence in creatures today.

It's important to distinguish Neander's producer-oriented teleosemantics from Millikan's (1984, 1989) well-known *consumer*-oriented view. Millikan holds that the content of a representation is, roughly, the external condition that historically needed to be present for the consumer of the representation to perform its proper function. In other words, it is the condition that obtained in those cases where the representation led its consumers to initiate behavior that contributed to reproductive success. For example, consider a representation that induces its consumer to initiate predator-evasion behavior. On Millikan's view, this representation represents *predator*, because this is the condition that needed to be present for evasion behavior to promote survival and reproduction. Call this the *success condition* for evasion behavior.

Critically, as others have pointed out, the success condition for the behaviors that a representation's consumers are induced to perform need not be conditions with which the organism is ever in causal contact. This creates *prima facie* difficulties for Millikan's view. Pietroski (1992) illustrates this with a fictional scenario: Kimus are an alien species whose only predators are snorfs. The kimus live near a hill, and the snorfs prowl the outskirts of the hill everyday at dawn. Fortuitously, some kimus develop a mechanism, M, which causes them to enter a type of brain state, B, in response to red objects. Kimus enjoy B-states, so those with mechanism M reliably move toward any red object in their vicinity. Now suppose that every morning at dawn, a red object appears at the top of the hill. Kimus with M will seek out this object, and this enables them to avoid predation. As a result, M spreads throughout the kimu population. As Pietroski notes, the *success condition* for the behaviors that B-states induce their consumers to perform has nothing to do with red things, and everything to do with the absence of snorfs. Accordingly, on Millikan's view, B-states represent *absence of snorfs*, or *fewer snorfs this way*, rather than *red*. This is so even though (we might imagine) snorfs and their properties are causally irrelevant to the production of B-states.

Against this, CT imposes a strict causal constraint on content determination. For R to pick out C, Rs must have in the past been in causal contact with Cs, since the system that produces Rs must have done so in response to Cs. Thus, Neander contends that CT delivers a different, and more plausible, verdict in the kimu case (2017: 134-136). Although B-states induce behavior that is adaptive given the absence of snorfs, the mechanism that produces B-states functions to produce them *in response to* red items, and not in response to the absence of snorfs. As such, CT generates the result that B-states represent *red*.

Neander notes that CT is a "first-pass" principle. She later enriches it in two ways. First, she requires that perceptual representations represent their contents in an analog format, which exploits principles of second-order similarity (2017: ch. 8). Second, she introduces a separate condition for ruling out inappropriately proximal contents—e.g., properties of retinal images as opposed to properties of external objects (2017: ch. 9). The first enrichment won't matter here, so I'll put it aside. Regarding the second, I'll simply grant for present purposes that Neander's strategy rules out proximal contents, so I'll put this aside too (although the strategy comes up briefly in section 6.3). The issues I'll discuss below will only involve competitions among distal contents.

Neander's primary arguments for CT rely on its putative fit with mainstream information-processing explanations in cognitive science. She appeals, for example, to the process of auditory localization from interaural time differences (2017: 28-29) and to an empirical case study of visual localization impairment (2017: 29-34; McCloskey & Rapp 2000) to argue that cognitive scientists routinely attribute malfunction-permitting information-processing functions to the systems they study. That is, such systems are described as having the function of processing certain information, even if they currently fail to perform that function.

Two virtues of CT should be emphasized. The first is one that the theory shares with other versions of teleosemantics. Namely, the appeal to proper functions offers a *prima facie* attractive way of understanding *mis*representation. If a system S has the proper function of producing Rs in response to Fs, then misrepresentation occurs when S produces an R in response to a non-F.[6] Second, by requiring that representations stand in causal relations to the properties that they represent, the view ensures that the ability to perceptually represent property F is grounded in past perceptual encounters with Fs. As we saw in Pietroski's example, views that lack any such causal component risk delivering dubious content ascriptions.[7]

In what follows, I'll be concerned with two questions regarding the foregoing account: (1) What are CT's implications for the rich/thin debate? and (2) Considering these implications, how well does CT really mesh with typical information-processing explanations in vision science? But to answer these questions, we'll first need a better handle on a key notion within CT—the notion of a C-type event causing a representation tokening *in virtue of* its C-ness. This is where I turn next.

3. Toads, Flies, and Causes

3.1. Vertical Problems and Property-Sensitive Causation

Every psychosemantic theory must address problems of content indeterminacy. Such problems are sometimes divided into "horizontal" problems and "vertical" problems (Godfrey-Smith 1989). Consider a representation R that is typically tokened in the presence of dogs. On such occasions, Rs are caused by a complex sequence of events involving the dog, the transmission of light through the air, the impingement of electromagnetic energy on the retina, and physiological processes in the brain. The *horizontal* problem is the problem of determining how "distal" R's content is. Should we interpret it as representing a property of the dog, or a property of the proximal stimulus? Suppose,

[6] There is a deeper disagreement within teleosemantics about *when* misrepresentation occurs. In earlier developments of her position, Neander explicitly claims that sensory-perceptual systems cannot misrepresent unless they *malfunction* (Neander 1995). Millikan's view, on the other hand, allows for pervasive misrepresentation by normally functioning systems (Millikan 1989).

[7] There are familiar problems with a causal requirement on perceptual content that I won't rehearse here. One issue is that certain perceptual systems may be subject to *normal misperception* (Matthen 1988; Mendelovici 2013; McLaughlin 2016). When they function normally, they tend to produce states that misrepresent their environment. Related problems afflict most naturalistic approaches to content determination, however, so they are at least not specific to the producer-teleosemantic approach.

however, that we solve the horizontal problem, and settle on a specific spatiotemporal stage in the causal chain as the one relevant to fixing R's content. The *vertical* problem is the problem of determining *which* property of the object(s) or event(s) at that stage R represents. Should we interpret R as representing *dog*, *animal*, or *dog-shaped object*?

The frog/toad example raises a vertical problem. When a fly buzzes through a toad's field of vision, the toad will typically snap its tongue at it. The mechanism underlying this capacity involves a pathway from the toad's retina to its optic tectum (Ewert 1987; Neander 2017: 100-109). The firing of T5(2) cells in the latter area causes "prey-seeking" behavior, such as orienting and tongue-flicking toward a moving object. But this mechanism will be activated in response to any stimulus with the appropriate size, shape, and motion properties. Specifically, T5(2) cells tend to fire in response to any small, dark, elongated object moving in a direction aligned with its axis of elongation.

There seem to be multiple reasonable ways of assigning content to the firing of T5(2) cells. These include *fly*; *toad food*; *moving toad food*; *small, dark, and moving*; etc.[8] Unless a psychosemantic theorist is comfortable with widespread content indeterminacy, she must say something about which if any of these attributions is correct. Neander (2017: ch. 5) argues that T5(2)-firing represents a certain low-level configuration of features—roughly, *small, dark, elongated, and moving in the direction of elongation*. I'll abbreviate this low-level configuration "SDM," for "small, dark, and moving."

Neander grants that the toad's visual system both has the function of detecting toad food and has the function of detecting SDM objects (2017: 160; see also Neander 1995: 114-115). The system has both tendencies, and both contributed to the survival and reproduction of its ancestors. Nevertheless, she claims that when flies caused the toad's visual system to produce T5(2)-firing in

[8] For rich interpretations of the toad, see Millikan (1986), Price (2001), and Shea (2007). For thin interpretations, see (aside from Neander) Fodor (1990) and Schulte (2017). Others, such as Papineau (1998), hold that the case is simply indeterminate.

function-conferring circumstances, they did so *in virtue of* being SDM, and not in virtue of being toad food:

> Consider past function-conferring selection. This involves past occasions on which the relevant pathways in the [toad]'s visual system contributed to [its] survival and reproductive success. On such occasions, the relevant state changes in the [toad]'s optic tectum were produced in response to visual targets in virtue of those targets being SDMs and *not* in virtue of their being [toad] food or flies. Causation is property-sensitive. (159)

Accordingly, given the "in virtue of" requirement in CT, T5(2)-firing represents visual targets as SDM, and not as flies or moving toad food.[9]

But why should we accept that when a fly buzzes past the toad, it causes T5(2)-firing in virtue of being SDM, and not in virtue of being toad food, or a fly? This claim is difficult to evaluate without a further account of *how* causation is property-sensitive. Suppose, for example, that the closest worlds in which the toad's visual targets are not edible are also worlds in which they are not SDM. Then it will typically be true that if a target had not been toad food, T5(2) cells wouldn't have fired. Thus, at least on a simple counterfactual analysis of causation, it is not *obvious* that the property of being toad food is causally irrelevant to T5(2)-firing. The claim requires positive support.

Neander (2017: 270-271) recognizes the need to say more. She suggests two ways of clarifying property-sensitive causation. The first appeals to Woodward's (2003) interventionist approach. On Woodward's theory, variable x causes variable y if, roughly, there are at least some conditions in which interventions on x are associated with changes in y. An intervention is understood as a "surgical" change to the value of x that does not directly alter any other variable in the system. Against this background, Neander proposes the following:

> (a) If the visual target were changed from being SDM to not being SDM (but nothing else were changed), then the [toad]'s visual system would change from producing Rs to not producing Rs.

[9] See Pautz (2013) for a parallel argument concerning human perception of high-level properties. Pautz claims that the visual system is sensitive to a tomato's low-level properties, but never to the property of being a tomato.

(b) If the visual target were changed from being a fly or [toad] food (F) to not being F (but nothing else were changed), then the [toad]'s visual system would not change from producing Rs to not producing Rs. (270)

Let "R" pick out T5(2)-firing. The idea is that if we intervened on whether the target was SDM, this would (in appropriate conditions) produce changes in the toad's response to the target, but if we instead intervened only on whether the target was nutritious for the toad, this would not produce changes in the toad's response to the target. Thus, Woodward's theory is supposed to deliver the result that the property of being SDM causes T5(2)-firing, while the property of being toad food does not.[10]

The problem with this approach is that while it may justify preferring SDM to toad food as the content of T5(2)-firing, it will not justify preferring SDM to other locally coinstantiated properties. Plausibly, a fly exhibits the appropriate motion pattern for eliciting T5(2)-firing *because* (and only while) it is in a type of internal physiological state I_1. Something similar is true of all the other edible critters that the toad encounters in its normal environment, although the specific internal states will vary. Now take the set of such states, $I_1 \ldots I_n$, and consider the property: being in at least one of $I_1 \ldots I_n$. If we change whether a critter exhibits this property, this will change whether the toad's visual system produces T5(2)-firing, because it will change whether the critter exhibits SDM. Accordingly, interventions on *either* property are associated with changes to the toad's response, so both remain candidate contents for T5(2)-firing.

Fortunately, Neander's second way of clarifying property-sensitive causation avoids this difficulty. She proposes the following:

c causes e in virtue of c being a C-type event rather than in virtue of c being a Q-type event if and only if:

[10] It's critical for this claim that interventions on whether something is nutritious for the toad do not *cause* changes in whether it is SDM. On Woodward's theory, when we intervene on x to determine whether it causes y, we do *not* hold fixed those variables that lie on the path from x to y. I am unsure how Neander envisages intervening on toad food. As such, it is not entirely clear to me that such interventions would *not* be associated with changes in whether the toad's visual targets are SDM, and so in turn with changes in T5(2)-firing. In any case, I won't pursue the matter here.

i. c causes e.
ii. had c instantiated C but not Q, e would have occurred, and
iii. had c instantiated Q but not C, e would not have occurred. (271)

Consider again the event of a fly buzzing past a toad, which causes a T5(2) cell to fire. Had this event involved an SDM target that wasn't toad food, the cell would still have fired. But had it involved toad food that wasn't SDM, the cell would not have fired. Likewise, had the event involved an SDM target that wasn't in any of $I_1 \ldots I_n$, the cell would still have fired. But had it involved a target in one of $I_1 \ldots I_n$ that wasn't SDM (e.g., because critical connections to motor effectors were severed), the cell would not have fired. Given CT, this seems to secure the result that T5(2)-firing represents SDM, and not toad food.

One concern is that this approach only specifies when an object or event causes a representation-tokening in virtue of one property *rather than* in virtue of another. So, problematically, it seems to imply that there are cases where an event c causes another event e in virtue of being C rather than in virtue of being Q, but also causes e in virtue of being D rather than in virtue of being C. This will be the case if (i) e would still have occurred had c been D but not C, (ii) e would not have occurred had c been C but not D, (iii) e would still have occurred had c been C but not Q, and (iv) e would not have occurred had c been Q but not C.

However, we can modify the account to avoid this difficulty. Say that a property F of object or event c is *screened off* with respect to an effect e just in case there is some other property G of c such that if c had instantiated G but not F, then e would still have occurred, but if c had instantiated F but not G, then e would not have occurred.[11] On this revised account, we can say that c causes e in virtue of being F just in case no other property of c screens off F with respect to e.

[11] This is essentially what Yablo (1997: 268) calls the *superproportionality* requirement on causation. Yablo notes that superproportionality faces deep problems of its own, which I won't try to address here. For example, the requirement gives rise to cases of "collective self-destruction" in which *no* property satisfies it. This will occur if for every property X of c, there is some other property of c that screens off X with respect to effect e. Suppose that we have candidate

I am unsure whether Neander would accept this strict screening-off principle of property-sensitive causation. However, in what follows I'll only suppose that it plays a *tie-breaking* role in her account. If we are unsure which of two properties is the one in virtue of which one event causes another, then we can appeal to screening-off to break the tie. This, in any event, is how Neander uses it.[12]

3.2. Why CT Rules Out High-Level Content

Given this approach to property-sensitive causation, CT has quite general implications for the contents of perceptual representations. Specifically, it entails that no high-level properties are perceptually represented—or at least that none are represented by the sorts of perceptual states that CT is meant to cover.[13] I'll illustrate this with the case of perceiving causation, and then draw out the general lesson.

Experimental work on the perception of causation provides, I believe, the most compelling case for high-level content in the visual system, so this will be my main example in what follows. The perception of causation is typically studied through variations on Albert Michotte's (1963) famous paradigm. One circle moves toward another circle, contacts it, and then the second starts moving. When the second circle moves immediately following contact, and in the same direction as the first, observers report a strong impression of causality.

---

properties of c: F, G, and H. It is possible for F to screen off G, which screens of H, which in turn screens off F. In this case, no property qualifies as the one in virtue of which c causes e.

[12] An immediate worry about the screening-off principle, as I've introduced it, is that it will systematically privilege proximal properties over distal properties—e.g., properties of the retinal stimulus over properties of external objects. However, it's important that the screening-off strategy is only used to resolve competitions between multiple properties of the *same* object or event. It is *not* supposed to resolve competitions between properties of distal and proximal stimuli. Neander proposes a separate strategy for avoiding inappropriately proximal contents (see section 6.3 below).

[13] Neander limits her account to "nonconceptual" representations, by which she seems to mean something like "pre-cognitive," or "deployable without cognitive involvement." As such, she might hold that perception sometimes deploys *conceptual* representations, presumably due to higher-level cognitive influence, and that these have high-level content. However, the putative cases of high-level content I'll be concerned with here are not plausibly construed as resulting from cognitive penetration. For example, the perceptual attunement to causation that I'll discuss presently is found in very young infants (Leslie & Keeble 1987; Saxe & Carey 2006; Kominsky et al. 2017), and can come apart from a perceiver's beliefs about whether a display is genuinely causal (Michotte 1963: 85; Scholl & Tremoulet 2000: 306; although see Rips 2011).

Causal events in Michotte displays can be classified into two types: (1) events in which the second circle moves at *less* than double the speed of the first, and (2) events in which the second circle moves at *more* than double the speed of the first. Call events of type 1 "launch-like" events, and events of type 2 "trigger-like" events. It would plausibly have been adaptive for our ancestors to be sensitive to the difference between launch-like and trigger-like events. It is a constraint of Newtonian mechanics that if one object contacts another, then, no matter their relative masses, the second can move at a speed at most double the speed of the first based on the force of their collision alone (Kominsky et al. 2017). If the second object moves at more than double the speed of the first, then there must have been a hidden force that contributed to its movement. So a perceptual boundary between launch-like and trigger-like events would be expected to do quite well in tracking the distinction between *launching*—an event-type in which one object causes another to move and the second's movement is solely a result of the force of their collision—and *triggering*—an event-type in which one object collides with and transfers force to another, prompting an independent causal source to propel the second in the same direction.

There is evidence that impressions of causality in Michotte displays are rooted in autonomous stages of visual processing. Furthermore, there is evidence that within the class of causal events, vision indeed draws a categorical boundary between launch-like and trigger-like events.

One way to support the claim that a capacity is genuinely perceptual is to show that it is susceptible to retinotopic adaptation aftereffects—that is, aftereffects specific to the retinotopic location in which an adaptation stimulus was presented. The reason is that even if cognitive representations undergo adaptation, there is no reason to expect their adaptation to be specific to a specific location on the *retina* (Gao & Scholl 2013; Block 2014). Rolfs et al. (2013) showed that adaptation to causal events is retinotopic, and that this adaptation is not easily explained by adaptation to simple spatiotemporal or kinematic features. Participants viewed a series of launch-like collisions, and then saw an ambiguous event in which the first circle partially or fully overlapped the second before the second began to move. In these ambiguous events, observers could either have an impression of causal launching, or of non-causal "passing," in which the first circle seemed to move continuously in front of the second. Rolfs et al. showed that for a given amount of spatial overlap, participants were more likely to report seeing a non-causal pass than a causal launch after viewing a series of launching events, and that this effect was significantly stronger when the ambiguous event was shown in the same retinotopic location as the adaptation stimuli. However, no retinotopically specific aftereffect on passing reports was found when participants were instead adapted to non-causal (but otherwise perceptually similar) events in which the first circle moved across the second and stopped on the other side before the second began to move.[14] A similar asymmetry between causal and non-causal events has also been recently documented using a continuous flash suppression paradigm (Moors et al. 2017).

Thus, there is evidence that vision draws a boundary between causal and non-causal Michotte displays. Moreover, a recent visual search study suggests that *within* the class of causal collisions, vision draws a boundary between launch-like and trigger-like events. Kominsky et al. (2017) showed subjects three events side-by-side and required them to indicate which was the "asymmetric" event—the one in which two objects moved at different speeds. Asymmetric events could either involve the first object moving at three times the speed of the second (3:1 events), or the second moving at three times the speed of the first (1:3 events). Kominsky et al. found that

[14] More recently, Kominsky and Scholl (2016) replicated the Rolfs et al. (2013) findings, but also found suggestive evidence that vision respects a categorical boundary between launching and triggering. After being exposed to a series of launching events in which the first object moved either much faster than or at the same speed as the second, participants were *not* more likely to see an ambiguous triggering event (in which the second object moved much faster than the first) as a non-causal pass. Conversely, adaptation to a series of triggering events did transfer to an ambiguous launching event, making it more likely to appear as a pass. Although this asymmetry is somewhat puzzling, the results provide evidence that adaptation more readily transfers within the launching or triggering categories than across them.

subjects were significantly faster to detect 1:3 events than 3:1 events under these conditions.[15] 1:3 events, but not 3:1 events, seem to "pop out" among symmetric (1:1 or 3:3) events. Why should this be? Note that 3:1 events and symmetric events *both* fall within the launching category, while 1:3 events fall within the triggering category. Thus, a natural explanation for the visual search patterns Kominsky et al. found is that perception categorized the causal displays in accordance with whether they were "launch-like" or "trigger-like," and that categorical differences facilitated pop-out.[16]

The empirical facts about how we perceive Michotte displays are by no means entirely settled (e.g., Rips 2011). But the evidence strongly suggests that there is a genuine form of perceptual categorization occurring here. Suppose, then, that visual representations R1 and R2 correspond to the "launch-like" and "trigger-like" categories, respectively. Our question is whether R1 and R2 have *genuinely causal* contents or not.

Consider R1. While launching is a constitutively causal event-type, the property of being a launching is coinstantiated with a configuration of lower-level features that don't involve causation. The following conditions are diagnostic of launch-events in ecologically normal environments:

(1) Object A moves in direction D at speed S.

(2) Object A contacts stationary object B at t.

(3) Immediately after t, object B moves in direction D at a speed no more than 2*S.

Given the evidence described above, R1 cannot plausibly be construed as representing any individual component of this configuration. But it remains possible that R1 represents the *whole* configuration (1)-(3). And, critically, it will *always* be possible to enrich this schema to produce a viable alternative to genuinely causal contents, no matter what additional experimental controls are

[15] This pattern only held for displays that could be interpreted as causal. If, for example, the three events in the search display were ones in which the first object failed to come into contact with the second, or there was a noticeable temporal gap between contact and movement of the second object, the difference in search time for 1:3 versus 3:1 events disappeared.

[16] I rely here on the conventional wisdom that pop-out effects are indicative of perception (Treisman & Gelade 1980).

investigated. Call the configuration (1)-(3) *launching*$_G$, for "launching Gestalt" (see Siegel & Byrne 2016).

CT decides this case in favor of the thin view. CT tells us that R1 represents the property C such that the visual system functions to produce R1s in response to C-type events *in virtue of* their C-ness. Given Neander's screening-off interpretation of property-sensitive causation, we get the result that, if R1 represents either, it represents launching$_G$, and not launching. If an event exhibited (1)-(3) without being a launching, R1 would still be tokened. However, if an event were a launching but did not exhibit (1)-(3), R1 wouldn't be tokened.[17]

The important lesson is this. High-level properties, if they are to be recovered, can only be recovered via their association with diagnostic lower-level properties.[18] Thus, it will, as a general rule, be possible to find a complex, potentially disjunctive category C involving these low-level properties such that instances of C that lack the relevant higher-level property H are perceptually indistinguishable from those that possess H. Therefore, if we hold C fixed while varying H, the perceptual system will continue to respond the same way. On the other hand, if we hold H fixed while varying C, the perceptual system will not continue to respond the same way. Therefore, if CT is correct, and property-sensitive causation is interpreted according to the screening-off principle, then high-level properties are not perceptually represented.[19]

---

[17] Regarding the second counterfactual, one might wonder whether it is even *possible* for an event to be a launching without exhibiting (1)-(3). Answer: It *is* possible if we consider a world in which the physical laws are different, allowing launched objects to move, say, four times faster than their launchers. However, even if it weren't possible, then the counterfactual would be satisfied vacuously. And Neander will need to rely on vacuously satisfied counterfactuals anyway. Consider T5(2)-firing again. Neander takes this state to represent SDM because this is ostensibly the property in virtue of which targets cause T5(2)-firing. But suppose we ask why SDM fails to be screened off by *moving*. The answer, presumably, is that (i) if a target were moving but not SDM, then T5(2) cells would not fire, and (ii) if a target were SDM but not moving, T5(2) cells would fire. But (ii) has a logically impossible antecedent, so can only be satisfied vacuously.

[18] Burge (2014: 575) also emphasizes this point. Vision scientists have devoted considerable attention to the diagnostic features of certain high-level categories. In the case of gist perception, vision scientists have attempted to specify the lower-level features diagnostic of particular scene categories (Oliva & Torralba 2006). Similarly for the detection of animals in cluttered displays (Delorme et al. 2010).

[19] Note that this is a significant departure from Millikan's teleosemantics. If we grant that the visual system tokens representations R1 and R2, it is plausible that their consuming systems put them to use in guiding behavior that is specifically adaptive in the presence of launching and triggering, respectively. After all, it seems likely that *these* are the

Neander (2017: 119-121) explicitly embraces this result as regards natural kind properties. She deems it implausible that perceptual systems are capable of representing high-level contents like H_2O (as opposed to XYZ), and acknowledges that, on her view, perceptual systems are limited to representing the "surface appearances of things" (120). However, it's important to note that CT rules out rich contents of *all* sorts—not merely natural kind properties.

4. Problems for CT

I'll now argue that Neander's theory faces problems, and that these problems trace directly to the feature that makes it thin-friendly. Specifically, they arise from the tight connection that the view imposes between representation and discriminability.

CT incorporates a notion of property-sensitive causation, and Neander's most promising way of interpreting this notion enlists a screening-off principle. The latter principle tells us that to settle whether a representation R represents C or Q, where C and Q are coinstantiated properties of the external event e that caused R, we ask the following: (i) Would R still have occurred had e been C but not Q? And (ii) Would R not have occurred had e been Q but not C? If the answer to both questions is "yes," then C screens off Q, and so beats it out for the distinction of being R's content. Note that this principle also enforces a strict discriminability requirement on representation: Suppose that all Cs are Qs, and that a perceptual system S cannot make any fine-grained discriminations within the class of Qs—i.e., S responds the same way to all Qs. Then Q will screen off C with respect to representations produced by S, because any variation in C that leaves Q fixed will not cause a change in which representations S produces.

We've seen that the screening-off principle rules out high-level content, and so supports the thin view. The problem is that it also leads to *low-level* content ascriptions that deviate significantly

categories that it is ultimately important for us to detect for purposes of survival and reproduction. As such, it is quite plausible that (e.g.) launching is the event-type that constitutes the normal condition for R1's consuming mechanism to perform its proper function. Thus, despite their common emphasis on proper functions, Neander and Millikan part company in a significant way concerning the contents of perceptual representations.

from those favored in mainstream vision science. This is so even if we confine our attention to early levels of visual processing, where a strict discriminability requirement would be most promising. Accordingly, the account is revisionist in a particularly unfortunate respect. Namely, it is inconsistent with the very kinds of scientific content ascriptions that Neander (2017: 98) suggests comprise the raw data for psychosemantic theorizing.

Consider motion detection (see Frisby & Stone 2010: ch. 14). Our visual system must recover motion on the basis of signals from a collection of photoreceptors. Most theorists agree that, to do this, vision relies heavily on a simple type of mechanism called a "Reichardt detector," which Burr and Thompson (2010: 1433) call "the backbone to any motion detector." Suppose that visual cells A and B are sensitive to nearby regions of space P1 and P2, respectively, and that both send signals to cell C when they detect a target. Suppose, further, that C fires just in case it receives simultaneous signals from A and B. The difficulty with getting C to indicate P1-P2 motion is that a target moving from P1 to P2 will not cause simultaneous signals to C, since A will send its signal before B does. However, suppose that we place a delay on A's input, so that when a target enters A's receptive field, it takes a bit longer to send a signal to C (see fig. 1). Then when a target moves at the appropriate speed from P1 to P2, C will receive concurrent signals from A and B, so it will fire. C thus achieves selectivity for motion at a certain speed from P1 to P2.[20]

Place figure 1 about here.

[20] The complete Reichardt detector combines this mechanism with a mirror-reversed one selective for P2-P1 motion, and the final output is produced by subtracting the second mechanism's outputs from those of the first. The arguments below are unaffected by this extra complexity, however, so I'll leave it aside.

There is evidence that motion perception in the fly is based primarily on Reichardt detectors (Haag et al. 2004), and it is widely thought that many areas of human motion processing work by means of systems of elaborated Reichardt detectors (van Santen & Sperling 1985; Lu & Sperling 2001). These are more complicated, but they operate on the same fundamental principle: a "delay and compare" process in which signals deriving from separate regions of space are compared after one of them passes through a delay filter.

Now consider a unit, D, that integrates information from several simpler Reichardt detectors so that it reliably responds to targets moving through regions P1-P6. How should we describe D's content? The first thing to ask is how vision scientists really do characterize the contents of such units. The point of these units, in standard information-processing models, is to pass on information about local *motion* (or at least directional components of motion). They are construed as representing something like *motion from P1 to P6 at speed s*. (Haag et al. (2004: 16337) simply refer to the Reichardt detector as a "motion detection algorithm.") But there is another option. We could insist on interpreting D as representing *something within region P1, then something within P2,…, then something within P6*. Call this *P1-P6 succession*, for short. Vision scientists opt for motion content over succession content. Presumably, the rationale is that although D is causally sensitive to succession, this is simply *how* it achieves causal sensitivity to motion, and there are independent reasons to favor motion content over succession content.

However, the screening-off principle entails that if D represents either, it represents succession, and not motion. Where c is an event involving P1-P6 motion, and so P1-P6 succession, had c instantiated the latter property without the former, D would still have fired, but had c instantiated the former without the latter, D would not have fired. (Assuming that P1-P6 motion metaphysically necessitates P1-P6 succession, the latter holds vacuously.) In case one is tempted to reject these empirical claims, note that it was one of the earliest discoveries in perception science that succession without motion fools visual motion-detectors. This is confirmable through any apparent motion display.[21]

We noted earlier that if high-level properties are to be perceived, they can only be perceived via their association with diagnostic lower-level features. But this type of situation is not unique to high-level properties. It also obtains for low-level properties like motion. For many properties F that vision scientists take the visual system to represent, F is associated with properties $G_1 \ldots G_n$ such that (i) $G_1 \ldots G_n$ probabilistically support, but don't necessitate, the presence of F, and (ii) if we held $G_1 \ldots G_n$ fixed while varying F, representation-producers in the visual system would continue to respond the same way. Neander's view eschews F-contents in all cases of this sort, so it departs significantly from mainstream vision science.

5. Fodor and Dretske

Theories that emphasize causal-informational relations in content determination are more likely than other views, such as Millikan's consumer teleosemantics or functional-role theories, to support thin content ascriptions (see note 19 above; see also Neander (2012: section 3.2)). In this section, I'll consider two more causal-informational theories, and show that while they are (or at least can be interpreted as) thin-friendly, they confront similar issues to those just raised for Neander's view.

On Fodor's (1990) asymmetric dependence theory, a mental representation R represents property F just in case: (i) it is a law that Fs cause Rs, and (ii) for all Xs that are not Fs, if it is a law that Xs cause Rs, then this law is asymmetrically dependent on the law that Fs cause Rs. The second condition means, roughly, that if we were to break the law linking Fs and Rs, we would also break the laws linking Xs and Rs, but for no X is it the case that by breaking the law linking Xs and Rs we would also break the law linking Fs and Rs.

[21] See also Barlow and Levick (1965) for a classic study verifying that early motion detectors in the rabbit will fire vigorously in response to appropriate sequences of flashes without genuine motion.

Fodor's view is thin-friendly.[22] For example, it is plausibly true that if we broke the law linking launching$_G$ with R1-tokening, we would also break the law linking launching with R1-tokening. But the converse is not (or at least not obviously) the case. More generally, laws relating high-level properties and perceptual representations are very likely to piggyback on laws relating diagnostic lower-level features and perceptual representations. Fodor's theory favors low-level contents in such cases.

However, recall our motion-sensitive cell, D. It is presumably a law (or at least follows from natural laws) that P1-P6 motion causes D-firing, and it is also a law that P1-P6 succession does. But the relationship between these laws is at best symmetric, and at worst asymmetric in the wrong direction. For if P1-P6 succession ceased to cause D-firing, then P1-P6 motion would as well. After all, D is sensitive to motion *by* being sensitive to succession, and it has no independent channel for gathering information about motion.

Dretske's (1988, 1995) indicator teleosemantics is in certain respects similar to Neander's account. Dretske (1995) holds that a perceptual system S represents a determinable property F just in case S has the biological function of indicating F. The way S performs this function is by entering specific states $s_1 \ldots s_n$ that indicate values $f_1 \ldots f_n$ of F. For example, a perceptual system dedicated to color perception might enter specific states that correspond to various determinate colors. Such states have, in Dretske's terms, *systemic* indicator functions—indicator functions derived from functions of the system of which they are states. Critically, Dretske understands indication in terms of probabilities, rather than causation. Roughly, x indicates y when $P(y|x)$ is sufficiently high. In Dretske's earlier writings (Dretske 1981), this value is set unrealistically to 1, while in his subsequent writings (Dretske 1986, 1988, 1995) no precise value is specified.[23]

---

[22] See Fodor (1990: 106-110) for discussion of the notorious bug-detector.

[23] There are obvious problems with placing a probability-1 requirement on indication if the notion of indication is to be at all useful in naturalizing mental representation. Since we are capable of misrepresentation, it would follow from this

Now consider the system—call it S—responsible for categorical perception of Michotte displays. The empirical evidence suggests that S can enter at least two types of states, s_1 and s_2, corresponding to the production of representations R1 and R2, respectively.[24] Thus, there are at least two interpretations of S. Either S functions to indicate various forms of *causation*, or S functions to indicate various forms of $causation_G$ (i.e., various causation-gestalts). Which option does Dretske's view favor? As stated, the view does not offer a clear answer, and Dretske does not address the issue directly. However, at least one of his strategies for resolving content indeterminacies clearly favors a thin interpretation. Unfortunately, the same strategy also entails the wrong interpretation of D (i.e., succession rather than motion).

If Dretske's theory is to favor thin ascriptions, it seems most likely to do so because of its indication requirement. After all, the theory's teleological component is quite unlikely to motivate $launching_G$ over launching (or SDM over toad food). Moreover, Dretske (1986) relies on an indication requirement for settling content indeterminacies in other cases—namely, indeterminacies between proximal and distal properties. His idea, roughly, is this: If a distal condition C can cause a representational state R to be tokened via two kinds of proximal input, f and g, then it's possible for $P(C|R)$ to be significantly higher than either $P(f|R)$ or $P(g|R)$, since C is present whenever R is tokened, but neither f nor g meets this condition. In this case, he suggests that we should favor C as the content of R.

Borrowing the same strategy, we *could* motivate the verdict that system S represents varieties of $causation_G$, rather than causation. The reason is simply that $P(launching_G|s_1) > P(launching|s_1)$,

requirement that mental representations typically do not indicate their representata. Dretske (1981) attempts to circumvent this difficulty by appealing to the *learning period* during which a mental representation is acquired. He holds that representations need only indicate their representata during their learning periods, and needn't indicate them afterward. (This proposal does not appear in Dretske's subsequent works.) However, as Fodor (1990) has argued, this won't do (see also Rupert 2008). First, there is no way to cleanly distinguish periods of concept learning from periods of concept mastery. Second, both learners and teachers make mistakes during periods of concept learning, so there is simply never a time in the concept acquisition process during which misrepresentation is impossible.

[24] Of course, if these representations simply *are* state-types, then s_1 = R1 and s_2 = R2.

and a similar asymmetry holds between triggering and its diagnostic gestalt. So S's states are in a sense "better" indicators of causation_G than causation. However—and this is the important point—this type of strategy would *also* recommend that D (or the state of producing D) represents succession rather than motion, because P(succession|D) > P(motion|D).

Thus, if Dretske's theory is thin, then this is plausibly because indication strength is used to resolve content indeterminacies. But if indication strength is used to resolve content indeterminacies, then, like Fodor's and Neander's accounts, the view delivers revisionist verdicts about low-level content.[25]

6. Two Remedies

We've considered three thin-friendly psychosemantic theories. In each case, we found that the element that makes the view thin-friendly also leads it to generate some striking departures from conventional content attributions in vision science. The current section returns to Neander's account. I'll scout two modifications that would improve the view's fit with mainstream vision science while retaining most of its key virtues. However, I'll also show that these remedies end up admitting high-level contents into perception. Moreover, while I'll discuss these issues within the context of Neander's account, I believe that their implications are more general. Any psychosemantic theory that adopts one of the two strategies discussed below for resolving vertical problems is likely to admit rich contents, because (given the empirical facts) it is likely to entail that we perceptually represent causal contents.

6.1. Representation-producing Response Functions

[25] Dretske (1990) offers a different strategy for handling certain indeterminacies. He considers cases where a state indicates both F and G, but indicates G only *by* indicating F. Dretske recommends the following treatment: "[W]hen an indicator, C, indicates both F and G, and its indication of G is *via* its indication of F…, then…C acquires the function of indicating that F" (826). Assuming that R1 indicates launching only by indicating launching_G, this strategy also supports the thin view. (Thanks to Adam Pautz for pointing this out.) Unfortunately, Dretske (1990) doesn't fully explicate what it means to indicate G by indicating F, so this proposal is difficult to assess (but see section 6.3 below). However, it is quite likely to confront a similar difficulty to those above. Plausibly, in whatever sense in which R1 indicates launching only by indicating launching_G, D indicates motion only by indicating succession.

Recall CT:

> CT: A sensory-perceptual representation, R, which is an (R-type) event in a sensory-perceptual system (S), has the content *there's C* if and only if S has the function to produce R-type events in response to C-type events (in virtue of their C-ness).

CT got into trouble because of the "in virtue of" criterion. That criterion plays a critical role, because it is needed for CT to have any chance of resolving vertical problems. But the way Neander understands it generates conflicts with mainstream vision science. Fortunately, however, I think that we can retain the core virtues of Neander's view while discarding the "in virtue of" criterion. This will enable us to explore alternative ways of resolving vertical problems.

The central idea behind CT is that content is grounded in response functions—more specifically, that it is grounded in functions to produce representations in response to external conditions. For clarity, I'll refer to this specific type of response function as a *representation-producing* response function. This is a *prima facie* attractive suggestion, because it provides for misrepresentation while also ensuring that representational content is grounded in prior perceptual encounters. But when exactly does a system have a representation-producing response function? Neander's own characterization of this notion relies on facts about property-sensitive causation. This is because property-sensitive causation is built into her idea of what it is for a representation to be produced *in response to* an external event. She writes: "To say that a system produced an R-type event in response to a C-type event is (here) to say that the C-type event, in virtue of its C-ness, caused the R-type event" (2017: 153). But this is just the requirement that I think we should avoid.

I'll work with a weaker notion of representation-producing response functions. This notion incorporates conditions that (I suspect) Neander would deem necessary but not sufficient for a system to have a representation-producing response function. I'll simply take them to be both necessary and sufficient. The idea is this: S has the function of producing R-type events in response to C-type events just in case S's ancestors produced R-type events, and in doing so met two conditions with respect to C-type events, which we can call *reliable causation* and *selective adaptiveness*.

If C-type events are a reliable cause of R-type events, then C-type events cause R-type events, and R-type events rarely occur in the absence of C-type events. (As above, I'll assume a broadly counterfactual notion of causation.) Thus, if our visual system has the function of producing R-states in response to red items, then it must be the case that red items caused the visual systems of our ancestors to produce R-states, and R-states were rarely produced in the absence of a red item.[26]

If producing R-type events in response to C-type events is *selectively adaptive*, then (i) R-production in the presence of C-type events is typically adaptive, and (ii) R-production in the absence of C-type events is (or would be) typically not adaptive. Thus, if the visual system has the proper function of producing R-states in response to red items, then it must be the case that, in the past, R-production in response to a red item tended to promote survival and reproduction, while R-production in response to a non-red item did not tend (or would not have tended) to promote survival and reproduction. Note that by "not adaptive" I do not mean *mal*adaptive. R-production in the absence of C needn't appreciably *harm* the organism's chances of survival or reproduction. It just doesn't improve these chances.

On this understanding of representation-producing response functions, it's clear that (e.g.) the toad's visual system has the function of producing T5(2)-firing both in response to toad-food events and in response to SDM events. Both are reliable causes of T5(2)-firing,[27] and both met the

[26] One might be tempted to weaken or drop the reliable causation constraint. Perhaps a system can have the function of producing Rs in response to Cs even if Cs *very rarely* cause Rs, while non-Cs frequently do. For example, perhaps a detector can have the function of firing in response to predators even if most of the things that cause it to fire are non-predators. This idea is particularly salient in consumer versions of teleosemantics (e.g., Millikan 1989: 283). I grant the force of the intuition. However, weakening the reliable causation constraint simply makes it *easier* to arrive at high-level content in perception, because our causal contact with low-level properties is typically more reliable than our causal contact with high-level properties. Thus, given my present dialectical aims, I'm not doing myself any favors by assuming the constraint.

[27] More specifically, both are reliable causes in the weak, counterfactual sense of causation. Consider cases in which a buzzing fly causes T5(2)-firing. It will typically be true in such cases that if an edible item *had not* been present, then

selective adaptiveness constraint. Likewise, early visual mechanisms have the function of producing D-firing both in response to motion events and in response to succession events. So vertical problems are as pressing as ever. The task, then, is to identify a further component that resolves these indeterminacies, and does so in a manner that comports with mainstream vision science. I'll now discuss two potential strategies that have some promise of achieving this result. I'll show that both end up admitting high-level properties into perceptual content.

6.2. Remedy 1: Informativeness

The first strategy appeals to an *informativeness* constraint on content determination.

Suppose that a system S has the function of producing R-type events in response to three properties or event-types, F, G, and H. One idea is that we should interpret R as representing *there's F* only if this content is *more informative* than either *there's G* or *there's H*. The rationale behind this norm is as follows. Perception aims to pass on representations to cognitive systems that are as informative about the environment as possible, as long as: (i) this level of informativeness does not significantly reduce its chances of accuracy (i.e., the probability that a representation's content is present, given that the representation is tokened), and (ii) this level of informativeness does not involve drawing finer distinctions than are relevant to the adaptive needs of the organism.

But when is one content more informative than another? This notion can be fleshed out using Dretske's (1981) notion of "nesting." In Dretske's analysis, the content *there's G* is nested in the content *there's F* just in case either (i) *there's F* logically entails *there's G*, or (ii) it follows from natural laws that (given ecologically normal background conditions) all Fs are Gs.[28] Given this, the

---

T5(2) cells would not have fired. Why? Because the nearest worlds in which an edible item is not present are unlikely to be ones in which some *other* (non-edible) SDM item is present. Rather, they are much more likely to be ones in which *no item at all* buzzes past the toad, in which case there is no viable stimulus for T5(2)-firing.

[28] Dretske (1981) does not include explicit reference to "normal background conditions," but a quick survey of his examples illustrates that he *must* assume it. For instance, Dretske (1981: 185) suggests that a sentence S of newspaper print nomically nests information about the people and events that it represents—call this information P. But clearly the relationship between P and S only holds relative to normal background conditions (roughly, conditions in which reporters are disposed to produce certain sentences of English in response to certain observations).

informativeness constraint says that if we are faced with a choice between *there's F* and *there's G*, and the former nests the latter but not vice versa, then we should select *there's F* over *there's G*. This comports with Dretske's own position about how to assign semantic contents in such cases (1981: ch. 7).

An appeal to informativeness resolves content indeterminacy in the motion case, and does so in the desirable way. Recall that cell D responds reliably to the successive occupancy of locations P1-P6. When the visual system is functioning normally, the following conditions hold. First, events involving motion from P1-P6 cause D to fire, and it is very rare for D to fire in the absence of motion from P1-P6. Second, D's firing in the presence of P1-P6 motion is typically adaptive, while D's firing in the absence of P1-P6 motion is almost never adaptive. Finally, it is necessary that all cases of P1-P6 motion are also cases in which locations P1-P6 are successively occupied, while the converse is not true. So motion content nests succession content, but not vice versa. The informativeness constraint thus recommends that D represents *P1-P6 motion*, rather than *P1-P6 succession*.

An informativeness constraint on content determination also has some precedent in the vision science literature. Scientific content ascriptions go hand-in-hand with computational-level theories of what a system is doing—the problem it is solving or the task it is performing (Marr 1982). Cognitive scientists adopt one ascription over another because they have a computational theory that mandates the ascription. Accordingly, if an informativeness constraint is built into one's computational theory, then it will also tend to govern ascriptions of content to the psychological or neurophysiological states to which that theory is applied.

Consider, then, Feldman's (2003, 2009) *minimal model theory* of how vision arrives at representations of the geometrical relations among edges. Feldman's approach is based on the important observation that certain relations in the 2-D retinal image are strongly predictive of corresponding relations in the 3-D environment (see also Biederman 1987; Albert & Hoffman 1995). This is the case with collinearity, parallelism, symmetry, and cotermination. For example, two elements that terminate at the same point in the retinal image are highly likely to have been projected by physical edges that terminate at the same point in 3-D space, and two edges that are collinear in the retinal image are highly likely to have been projected by collinear edges in 3-D space.[29] Feldman calls such special relations preserved under projection *regular* relations.

Many regular relations are special cases of one another. For example, collinearity is a special case of (and thus nests) parallelism, which is itself a special case of the relation two edges stand in just in case they are both located in space—the 'general position' relation. Accordingly, representing two edges as collinear is more informative than representing them as parallel, and representing them as parallel is more informative than representing them as in general position. Feldman notes that edge relations can be arranged into a partial order, where those lower in the order are special cases of those higher up (fig. 2). Given a pair of collinear and coterminating retinal image elements, then, the visual system needs to decide what relation to represent in the 3-D environment. Feldman recommends a very simple rule: Given image I, "choose the lowest interpretation in the partial order consistent with I" (2009: 880). In other words, given multiple interpretations compatible with the image, choose that interpretation which is *strongest* or *most informative*.

Place figure 2 about here.

[29] For two non-collinear edges to project collinear retinal images, the perceiver would need to occupy a very special viewpoint in which the edges just happened to be aligned from her perspective. The slightest movement would reveal their non-collinearity. Collinear edges, on the other hand, project collinear retinal images for every viewpoint in which both are visible.

I've suggested that, in practice, vision scientists posit solutions to vertical problems because these solutions are mandated by the computational theories that they endorse. Feldman's theory mandates ascribing contents according to an informativeness constraint. Consider a state-type R that is typically produced as a result of 2-D collinear elements on the retina, and only in such cases. We want to know whether R represents 3-D collinearity, parallelism, or general position. Problem: All of these relations are reliably instantiated by the physical edges that typically cause Rs. However, if we accept the minimal model theory, we resolve the ambiguity. We interpret R as representing collinearity because this is the lowest relation in the partial order—and so the most informative content—with which R is in reliable causal contact.

An informativeness constraint thus seems to deliver plausible, scientifically respectable solutions to certain vertical problems. It's important to underscore that the motivation for this constraint derived solely from considerations about *low-level* content. But if informativeness is really a viable principle for resolving vertical problems, this creates a problem for the thin-friendly psychosemanticist. For the same principle can also be used to argue that certain perceptual states have high-level content.

Recall representation R1, which is reliably produced in response to launch-like but not trigger-like events. It's plausible that the vast majority of external events that instantiate launching$_G$ are also genuine launchings. Thus, assuming that launching$_G$-events reliably cause R1-tokening, launching-events likely do so as well. Furthermore, R1-tokening in response to launching is typically adaptive, while R1-tokening in the absence of launching is typically not adaptive. Finally, launching content is more informative than launching$_G$ content, because it is a natural law that all launching-events (or at least all launchings in normal macroscopic environments) are launching$_G$-events, but it is not a law that all launching$_G$-events are launchings. So the content *there's launching* nests the content *there's launching*$_G$, and not vice versa. Thus, just as informativeness favored motion over succession, it seems to favor launching over its low-level diagnostic configuration.

To sum up: one way to bring Neander's account into conformity with content ascriptions in mainstream vision science is to require that perceptual states represent the most informative contents viable in the circumstances (i.e., given facts about causation and adaptiveness). However, if we adopt this route, then it becomes quite plausible that certain perceptual states have high-level contents—namely, causal contents.

More generally, any informativeness constraint on content determination inspired by Dretske's (1981) nesting rules is likely to deliver rich content in certain cases. Specifically, this is likely to occur whenever a perceptual representation reliably tracks both a high-level property F and a diagnostic low-level property G, and there are natural laws linking F and G, creating an asymmetric nesting relation between them. Thus, psychosemantic theories that prioritize informativeness are prone to entail that there are at least some perceptual representations of high-level properties.

6.3. Remedy 2: Hierarchy of Functions

A second strategy for resolving vertical problems could exploit the fact that perceptual systems seem to perform some representation-producing response functions *by* performing others, but not vice versa. For example, it seems right to say that early visual motion detectors function to respond to both succession and motion, but respond to motion *by* responding to succession. Conversely, they do not respond to succession by responding to motion. This hierarchical relationship is by no means unique to response functions. The heart performs its function of circulating oxygen by performing its function of pumping blood.

Interestingly, Neander (2017: 221-223) exploits hierarchical relations among response functions in addressing *horizontal* indeterminacy problems. She grants that a perceptual system may have representation-producing response functions vis-à-vis both a distal property and a proximal property of sensory input. However, she claims that in such cases, the system responds to distal properties *by* responding to proximal properties. Equivalently, the system responds to proximal properties *as a means to* responding to distal properties. She recommends opting for distal content in such cases. But Neander does not use this strategy for settling competitions among coinstantiated properties—e.g., between succession and motion, or SDM and toad food. She intends the strategy to apply only after vertical problems have been resolved through principles of property-sensitive causation. However, one might try to use it to settle vertical problems as well.

For present purposes, the important point is that *if* a principle of this sort recommends motion content over succession content, then it is quite likely to *also* recommend launching content over launching$_G$ content. I can envisage two ways of motivating the claim that the visual system responds to succession as a means to responding to motion. Both carry over seamlessly to the launching case.

The first way appeals directly to facts about adaptiveness. Historically, visual responses to succession were only adaptive when succession was coinstantiated with motion, so the visual system, it might be said, only "cares" about succession because it reliably signals motion. This provides a compelling reason for preferring motion to succession in our content ascriptions. However, if this consideration favors motion over succession, then a parallel consideration should favor launching over launching$_G$. The visual system only "cares" about launching$_G$ configurations because they reliably signal launching. Rare, coincidental cases where the relationship between launching$_G$ and launching is broken (e.g., where object A is a material ball, and object B is a patch of light) were not plausibly adaptive for the visual system to detect.

A second way of arguing that the visual system responds to succession as a means to responding to motion would appeal to causal asymmetries revealed by screening-off relations. The visual system reliably responds to both motion and succession, but succession screens off motion with respect to the activation of visual motion detectors. Thus, succession is to this extent more causally "immediate" than motion in eliciting this activation. While it's difficult to see how this fact could be called on to *favor* motion over succession in a theory of content determination, suffice it to note that if it did, then it would also favor launching. As discussed earlier, launching$_G$ screens off launching, but not vice versa.

More generally, any strategy for favoring motion over succession that exploits hierarchical relationships among functions is quite likely to also favor rich contents over thin contents in certain cases. For it is likely that perceptual systems sometimes respond to low-level properties as a means to responding to high-level properties. This will be plausible whenever (i) a perceptual state-type reliably tracks both a high-level property and a diagnostic low-level property, (ii) the low-level property causally screens off the high-level property with respect to the perceptual state-type, and (iii) the high-level property is more directly relevant to the organism's survival and reproduction.

Thus, both strategies for recovering scientifically respectable verdicts about low-level motion content also seem to favor ascribing high-level content in the case of perceiving causation. I draw two conclusions from this: First, it is remarkably difficult to identify a principle of content determination that categorically rules out high-level content without forcing revisionist verdicts about low-level content. Second, there are independently compelling principles of content determination that resolve indeterminacies about low-level content in the desirable way, and these principles plainly support the rich view over the thin view. While these aren't decisive arguments for the rich view, they provide clear marks in its support, and they should make thin theorists uneasy.

7. Generalizing the Challenge

But that's not all. A more general challenge lurks for anyone seeking a psychosemantic theory that categorically supports the thin view. If the content ascriptions favored in mainstream vision science are correct, then the visual system is at minimum capable of the following achievements:

(a) Where all Fs are Gs, visual states can determinately represent F even though their producing mechanisms have no ability to discriminate Fs from non-Fs within the class of Gs.

Example: Certain states of early vision represent motion, despite having no ability to distinguish this from mere succession.

(b) Visual states can determinately represent F even though F is only probabilistically, not deterministically, supported by information available in proximal stimulation.

Example: Visual states can determinately represent an object's luminance, even though this is only probabilistically, not deterministically, supported by the cues available in retinal input.

(c) Visual states can determinately represent F even though producing a representation of F requires first producing representations of the diagnostic features of F.

Example: Visual states can determinately represent volumetric shape properties, but such representations are produced through a series of computational steps that first produce representations of the geometrical properties of visible surface regions.

(a)-(c) mark respects in which the visual system's causal sensitivity to the properties that it represents is either imperfect or otherwise indirect. The important point is that (a)-(c) are supportable purely by appeal to uncontroversial ascriptions of *low-level* content. Accordingly, for a psychosemantic theory to rule out high-level content while maintaining consistency with these ascriptions, it must *not* do so by way of principles that entail that one or more of (a)-(c) is impossible. We saw above that Neander's screening-off principle rules out cases of type (a), and this creates trouble.

For a theory of content determination to support the thin view, it needs to identify some defect in our (or, better, the visual system's) relation to high-level properties, and show that this defect precludes their being perceptually represented. The problem is that the obvious defects one might point to are already present in our relations to many low-level properties. Although we lack perfect discriminatory abilities with respect to high-level properties, we also lack them with respect to low-level properties. And although high-level properties are only recoverable through "risky" computational transitions (ones resembling inductive inference, rather than deduction), the same is true of many low-level properties.[30]

In my view, the most promising route for a thin-friendly psychosemantics would be to devise an improved principle of property-sensitive causation. Perhaps it can be shown that the only distal properties causally relevant to perceptual representation-tokening are spatial properties, luminance, color, motion, and so on. This principle could be combined with a causal constraint on content determination to rule out high-level content. But such a principle would need to arrange things *just right* for the thin theorist to avoid significant conflicts with mainstream vision science. Specifically, it would need to rule *out* causation, animacy, numerosity, and so on, while ruling *in* motion—not to mention complex, partially occluded volume shapes.

8. Conclusion

Psychosemantics offers an important source of evidence in the debate between rich and thin views of perceptual content. But will a viable psychosemantics of perception support the rich view or the thin view? Neander's (2013, 2017) theory recommends the latter, and it is the most refined thin-friendly psychosemantic theory currently on offer. However, I've argued that the view faces difficulties, and that these difficulties originate from the very feature that makes it thin-friendly. I've also argued that this sort of issue is not unique to Neander's account, and I've shown that two natural, independently plausible remedies for the difficulty end up inviting high-level properties into perceptual content. Finally, I've argued that the key achievements needed for recovering high-level properties are already involved in recovering low-level properties, and that the visual system's discriminatory abilities are similarly imperfect in both cases.

[30] Gao and Scholl (2013: 221-222) have also emphasized this point.

I conclude that a viable psychosemantics of perception is quite unlikely to support the thin view. Recent work in vision science suggests that the visual system is systematically attuned to certain high-level properties. The obvious strategies for excluding high-level content in these cases are unsuccessful, because they also generate dubious verdicts about low-level content. Furthermore, there are independently compelling principles of content determination that, together with recent vision science, plainly favor the rich view. Finally, the obvious defects in our relation to high-level properties are also present in our relation to many low-level properties. Since these defects don't prevent us from perceiving low-level properties, they shouldn't prevent us from perceiving high-level properties either.

Appendix: Should We Prefer a Thin View of Toad Vision?

In this appendix, I'll consider two of Neander's arguments *against* ascribing high-level content in the case of toad vision. These arguments don't entail or rely on CT, but the principle is intended to cohere with them. I contend that these arguments are unsuccessful. *A fortiori*, they fail to motivate a thin view of perceptual representation more generally.

Discriminatory Abilities

In addition to representing targets as exhibiting a certain property or attribute (e.g., toad food or SDM), T5(2) cells also represent their targets as occupying certain locations. Location information is used to guide the toad's orienting and tongue-flicking behavior. Neander (2017: ch. 5) contends that the considerations that lead us to favor certain ascriptions of location content over others should also lead us to favor SDM over toad food.

T5(2) cells have spatial receptive fields. They respond to preferred stimuli within certain spatial regions, but not outside of these regions. So consider a particular T5(2) cell, T. Call the spatial region within which targets cause T to respond "C2." Let C1 be a region slightly smaller than C2, and C3 a region slightly larger than C2. Suppose that T does not exhibit finer sensitivity within C2, but rather responds uniformly to any SDM target within the region. If a fly in C1 causes T to fire, should we say that T's firing represents the fly as in region C1, C2, or C3? Neander answers (and I agree) that the best answer is obviously C2.

Neander takes this to underscore a general requirement according to which the contents of perceptual representations are closely tied to discriminatory abilities. Call the type of state in which cell T fires "W*." Neander writes:

> Notice that [these points] concern the *normal causal sensitivities* of the cells. Such normal causal sensitivities seem directly relevant to the localization content of the W*-type representations. Notice too that, if the processing pathways that produce these state changes *cannot* (normally) make more fine-grained discriminations among locations, then it cannot be their function to do so. (…) Therefore, these considerations suggest that the *representational content*, and not only the *informational content*, with respect to location, is not especially precise. (112)

In other words, because T can discriminate targets in C2 from those outside C2, but can't make finer discriminations within C2, it can't have the function of making such finer discriminations. Accordingly, we should say that its firing represents C2 and not, say, C1. But because T can discriminate targets in C2 from those in the non-C2 regions of C3, we can justifiably conclude that T represents C2 and not C3. The application to the SDM/toad-food case is straightforward. Because T5(2) cells cannot make finer discriminations within the class of SDM things, it cannot be their function to do so. Accordingly, we should say that T5(2)-firing represents SDM, and not toad food.

The problem with this argument is that the proposed link between discriminatory abilities and content is dubious. We can grant that T5(2) cells don't have the function of discriminating toad-food targets from those non-toad-food targets that are *also SDM*. (Something can't have the function of x-ing if it is wholly incapable of x-ing.) But why should we move from this to the conclusion that T5(2)-firing doesn't represent toad food? After all, we don't typically require that representing systems be capable of discriminating their representata from *all* look-alikes or decoys. I can think about Sam despite lacking a foolproof procedure for distinguishing him from his identical twin.

Elsewhere, Neander writes: "[T]he perceptual capacity to be explained is a capacity to distinguish between items in worm-like motion [i.e., SDM] and items that are not in worm-like motion. (…) In contrast, the normal toad has *no* capacity to distinguish nutritional from non-nutritional substances and so this is not the capacity to be explained" (2013: 31). But this last claim can be resisted. A defender of the toad food interpretation can hold that the toad *does* have a capacity for distinguishing nutritional from non-nutritional things. It's just that, like my capacity for distinguishing Sam from other people, this capacity is *imperfect*. The toad (like me) can be fooled.

Moreover, while content indeed closely coincides with discriminability in the localization case, there are differences between this and the SDM/toad-food case that may be relevant to content determination. First, while targets have a uniform probability of occupying any particular location within C2, SDM targets are far more likely to be toad food than not. Second, while it is typically adaptive for cell T to fire in response to targets anywhere within C2, it is not adaptive for T to respond to SDM targets that aren't toad food. Thus, subregions within C2 are uniform both in their probability of containing a target and in the adaptiveness of responding to targets so located. This is not true of all sub-categories within SDM.

Therefore, it's open to a defender of the toad-food interpretation to maintain that perceptual states can represent a property F despite an imperfect ability to discriminate Fs from non-Fs. There are at least two non-exclusive ways one might support such an interpretation. Either the system *very rarely* encounters those non-Fs that it can't discriminate from Fs, or it gains *no adaptive benefit* from tokening the relevant state-type in response to such non-Fs.[31]

---

[31] Neander (2013) agrees that recognitional concepts, like my abovementioned concept of Sam, need not be associated with perfect discriminatory abilities vis-à-vis their representata. However, she insists that a strict discriminability requirement does hold for sensory-perceptual representations. I see no compelling reason to accept the latter claim.

The Visibility Principle

Neander also argues that ascribing the content *toad food* to T5(2)-firing would conflict with a "central tenet" of information-processing explanations. She writes:

> [W]e can add to the idea that there should be a selective response to what is represented one of the central tenets of the information-processing approach to explaining vision—that visual content must be extracted from the retinas by subsequent processing. (…) In his popular text, *Vision Science*, Palmer articulates a closely related principle. As he puts it, it is a general tenet of contemporary information-processing accounts of vision that *visible properties must be represented before invisible ones.* (…) [F]or example, the surface of a cow that faces the viewer and is not occluded from the viewer's line of sight must be represented before subsequent processing represents the cow's far side, its insides, its canonical 3D shape or its being a cow as such. (116-117)

Although the visible/invisible distinction is somewhat obscure, I will pass over this complication. Presumably, by "visible properties," Neander has in mind properties that are directly relevant to how an object interacts with incident illumination and supplies light to the retinas. Other properties are either invisible due to occlusion or, as in the case of natural kind properties, have only an indirect relation to how the object interacts with light. The argument, then, is as follows.

(1) To represent invisible properties, a visual system must first represent certain visible properties.
(2) The property of being toad food is invisible.
(3) Therefore, to represent the property of being toad food, the toad's visual system must first represent certain visible properties.
(4) If the toad-food interpretation of T5(2)-firing is correct, then the toad's visual system does not represent any visible properties prior to representing toad food.
(5) Therefore, the toad-food interpretation of T5(2)-firing is incorrect.

However, premises 1 and 4 of this argument are questionable.

Start with premise 4. Against this, the toad-food interpretation does *not* obviously entail that toad food is recovered from retinal inputs with *no* mediating computational steps. Consider how T5(2)-firing is produced (Neander 2017: 106-109; Ewert 1987: 353-356). T5(2) cells receive excitatory inputs from other tectal cells—T5(1) cells—and inhibitory inputs from pretectal cells—TH3 cells. The former are selective for small objects extended parallel to their direction of motion, while the latter are selective for large square objects. If we interpret these as representing visible surface features, then the visibility principle is already satisfied. Even if T5(2)-firing represents invisible properties, it does so following inputs from cells that represent visible properties.

Turn to premise 1. This claim is touted as a "general tenet" of visual information-processing models. Nevertheless, while it is indeed *generally* true that visible properties come before invisible properties during visual processing, perhaps it is not *universally* true. Perhaps there are cases in which invisible properties are easily and reliably recovered from proximal stimulation, and it is possible in these cases to bypass the recovery of visible surface properties. The reason why vision *must* go through several stages of visible surface properties before representing something as a cow is that reliably detecting cowhood requires sensitivity to a complex pattern of features distributed across space. Because cells in early vision can't perform this kind of global computation, they can't reliably detect the property of being a cow. But the property of being toad food may be different. Perhaps this property *is* reliably detectable on the basis of a small number of local cues that can be integrated in early stages of vision. If so, then the case might be treated as an exception to Palmer's principle.

Finally, even if the visibility principle did provide adequate motivation for a low-level interpretation of T5(2)-firing, its application beyond this case is quite limited. Most who support rich content in perception will freely grant that high-level properties are recovered only on the basis of recovering low-level properties like shape, color, and motion. This would almost certainly be the case for, say, the perception of causation, facial expressions, or agency. But as we saw above, Neander's view entails that *no* high-level properties are perceptually represented—not merely that we can't represent them without first representing low-level properties.

Thus, neither discriminability considerations nor the visibility principle provides strong reason to confine the toad's (or our) visual system to low-level content.

References

Albert, M., & Hoffman, D. D. (1995). Genericity in spatial vision. In D. Luce, K. Romney, D. D. Hoffman, & M. D'Zmura (*eds.*), *Geometric Representations of Perceptual Phenomena: Articles in Honor of Tarow Indow on his 70th Birthday*. New York: Erlbaum, pp. 95-112.

Barlow, H. B., & Levick, W. R. (1965). The mechanism of directionally selective units in rabbit's retina. *The Journal of Physiology*, *178*(3), 477-504.

Bayne, T. (2016). Gist! *Proceedings of the Aristotelian Society*, *116*(2), 107-126.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, *94*(2), 115-147.

Block, N. (2014). Seeing-as in the light of vision science. *Philosophy and Phenomenological Research*, *89*(3), 560-572.

Block, N. (ms.) A joint in nature between cognition and perception without modularity of mind.

Briscoe, R. (2015). Cognitive penetration and the reach of phenomenal content. In J. Zeimbekis & A. Raftopoulos (*eds.*), *The Cognitive Penetrability of Perception*. Oxford: Oxford University Press, pp. 174-199.

Burge, T. (2014). Reply to Block: Adaptation and the upper border of perception. *Philosophy and Phenomenological Research*, *89*(3), 573-583.

Burr, D., & Thompson, P. (2011). Motion psychophysics: 1985–2010. *Vision Research*, *51*(13), 1431-1456.

Butler, A., Oruc, I., Fox, C. J., & Barton, J. J. (2008). Factors contributing to the adaptation aftereffects of facial expression. *Brain Research*, *1191*, 116-126.

Byrne, A. (2009). Experience and content. *The Philosophical Quarterly*, *59*(236), 429-451.

Delorme, A., Richard, G., & Fabre-Thorpe, M. (2010). Key visual features for rapid categorization of animals in natural scenes. *Frontiers in Psychology*, *1*, 1-13.

Dretske, F. I. (1981). *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.

Dretske, F. I. (1986). Misrepresentation. In R. Bogdan (*ed.*), *Belief: Form, Content, and Function*. Oxford: Oxford University Press, pp., 17-36.

Dretske, F. I. (1988). *Explaining Behavior: Reasons in a World of Causes*. Cambridge, MA: MIT Press.

Dretske, F. I. (1990). Reply to reviewers. *Philosophy and Phenomenological Research*, *50*(4), 819-839.

Dretske, F. I. (1995). *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Ewert, J. P. (1987). Neuroethology of releasing mechanisms: prey-catching in toads. *Behavioral and Brain Sciences*, *10*(3), 337-368.

Feldman, J. (2003). Perceptual grouping by selection of a logically minimal model. *International Journal of Computer Vision*, *55*(1), 5-25.

Feldman, J. (2009). Bayes and the simplicity principle in perception. *Psychological Review*, *116*(4), 875-887.

Fish, W. (2013). High-level properties and visual experience. *Philosophical Studies*, *162*(1), 43-55.

Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Fodor, J. A. (1990). *A Theory of Content*. Cambridge, MA: MIT Press.

Frisby, J. P., & Stone, J. V. (2010). *Seeing: The Computational Approach to Biological Vision*. Cambridge, MA: MIT Press.

Godfrey-Smith, P. (1989). Misinformation. *Canadian Journal of Philosophy*, *19*(4), 533-550.

Godfrey-Smith, P. (2013). Signals, icons, and beliefs. In D. Ryder. J. Kingsbury, & K. Williford (*eds.*), *Millikan and Her Critics*. Hoboken, NJ: Wiley & Sons Inc., pp. 41-62.

Haag, J., Denk, W., & Borst, A. (2004). Fly motion vision is based on Reichardt detectors regardless of the signal-to-noise ratio. *Proceedings of the National Academy of Sciences*, *101*(46), 16333-16338.

Kominsky, J., & Scholl, B. (2016). Retinotopic adaptation reveals multiple distinct categories of causal perception. *Journal of Vision*, *16*(12), 333.
Kominsky, J. F., Strickland, B., Wertz, A. E., Elsner, C., Wynn, K., & Keil, F. C. (2017). Categories and constraints in causal perception. *Psychological Science*, *28*(11), 1649-1662.
Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*(3), 265-288.
Lu, Z. L., & Sperling, G. (2001). Three-systems theory of human visual motion perception: Review and update. *Journal of the Optical Society of America A*, *18*(9), 2331-2370.
Mandelbaum, E. (2017). Seeing and conceptualizing: Modularity and the shallow contents of perception. *Philosophy and Phenomenological Research*. DOI: 10.1111/phpr.12368.
Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco, CA: W.H. Freeman.
Matthen, M. (1988). Biological functions and perceptual content. *The Journal of Philosophy*, *85*(1), 5-27.
McCloskey, M., & Rapp, B. (2000). Attention-referenced visual representations: Evidence from impaired visual localization. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(3), 917-933.
McLaughlin, B. P. (2016). The Skewed View from Here: Normal Geometrical Misperception. *Philosophical Topics*, *44*(2), 231-299.
Mendelovici, A. (2013). Reliable misrepresentation and tracking theories of mental representation. *Philosophical Studies*, *165*(2), 421-443.
Michotte, A. (1963). *The Perception of Causality*. Translated by T. R. Miles & E. Miles. London: Methuen & Co. (Original published in 1946).
Millikan, R. G. (1984). *Language, thought, and other biological categories: New foundations for realism*. Cambridge, MA: MIT Press.
Millikan, R. G. (1986). Thoughts without laws; cognitive science with content. *The Philosophical Review*, *95*(1), 47-80.
Millikan, R. G. (1989). Biosemantics. *The Journal of Philosophy*, *86*(6), 281-297.
Moors, P., Wagemans, J., & de-Wit, L. (2017). Causal events enter awareness faster than non-causal events. *PeerJ*, e:2932.
Neander, K. (1995). Misrepresenting & malfunctioning. *Philosophical Studies*, *79*(2), 109-141.
Neander, K. (2012). Teleological theories of mental content. In E. N. Zalta (*ed.*), *The Stanford Encyclopedia of Philosophy* (Spring 2012 Edition).
Neander, K. (2013). Toward an informational teleosemantics. In D. Ryder. J. Kingsbury, & K. Williford (*eds.*), *Millikan and Her Critics*. Hoboken, NJ: Wiley & Sons Inc., pp. 21-40.
Neander, K. (2017). *A Mark of the Mental: In Defense of Informational Teleosemantics*. Cambridge, MA: MIT Press.
Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, *155*, 23-36.
Papineau, D. (1998). Teleosemantics and indeterminacy. *Australasian Journal of Philosophy*, *76*(1), 1-14.
Pautz, A. (2013). Does the rich content view of experience matter? Comments on Matthew Frise. Talk delivered at the 2013 Meeting of the Pacific APA. Retrieved from URL: http://www.i-m.mx/apautz/adampautz/
Pietroski, P. M. (1992). Intentionality and teleological error. *Pacific Philosophical Quarterly*, *73*(3), 267-282.
Price, C. (2001). *Functions in mind: A theory of intentional content*. Oxford: Oxford University Press.
Prinz, J. (2013). Siegel's get rich quick scheme. *Philosophical Studies*, *163*(3), 827-835.
Rips, L. J. (2011). Causation from perception. *Perspectives on Psychological Science*, *6*(1), 77-97.

Rolfs, M., Dambacher, M., & Cavanagh, P. (2013). Visual adaptation of the perception of causality. *Current Biology*, *23*(3), 250-254.
Rupert, R. D. (2008). Causal theories of mental content. *Philosophy Compass*, *3*(2), 353-380.
Saxe, R., & Carey, S. (2006). The perception of causality in infancy. *Acta Psychologica*, *123*, 144-165.
Scholl, B. J., & Gao, T. (2013). Perceiving animacy and intentionality: Visual processing or higher-level judgment? In M. D. Rutherford & V. A. Kuhlmeier (*eds.*), *Social Perception: Detection and Interpretation of Animacy, Agency, and Intention*. Cambridge, MA: MIT Press.
Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299-309.
Schulte, P. (2017). Perceiving the world outside: How to solve the distality problem for informational teleosemantics. *The Philosophical Quarterly*. DOI: pqx052.
Shea, N. (2007). Consumers need information: supplementing teleosemantics with an input condition. *Philosophy and Phenomenological Research*, *75*(2), 404-435.
Siegel, S. (2009). The visual experience of causation. *The Philosophical Quarterly*, *59*(236), 519-540.
Siegel, S. (2010). *The Contents of Visual Experience*. Oxford: Oxford University Press.
Siegel, S., & Byrne, A. (2016). Rich or thin? In B. Nanay (*ed.*), *Contemporary Debates in the Philosophy of Perception*. New York: Routledge.
Susilo, T., McKone, E., & Edwards, M. (2010). Solving the upside-down puzzle: Why do upright and inverted face aftereffects look alike? *Journal of Vision*, *10*(13):1, 1-16.
Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97-136.
van Santen, J. P., & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A*, *2*(2), 300-321.
Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.
Yablo, S. (1997). Wide causation. *Philosophical Perspectives*, *11*, 251-281.

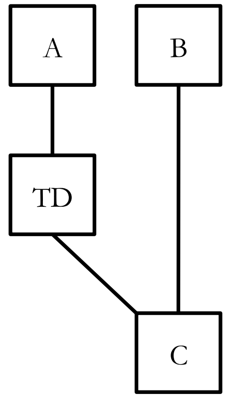

Figure 1. A simple Reichardt detector. Cells A and B are sensitive to nearby regions P1 and P2. TD is a temporal delay filter. C implements an AND-gate.

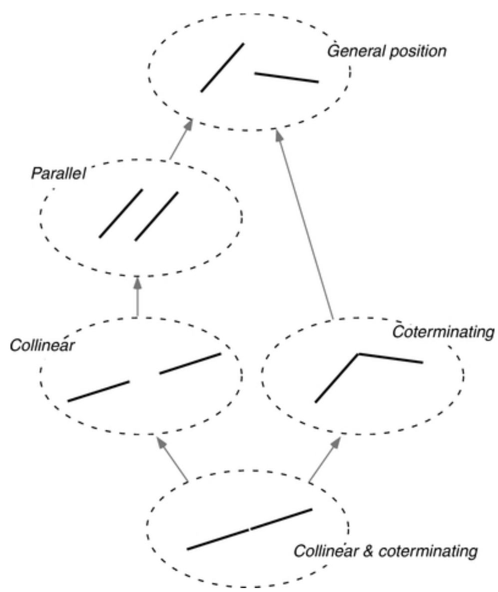

Figure 2. A partial order on geometrical relations between edge segments. *Source:* Feldman (2009: 877). Reprinted with permission.