

# Naturalizing intentionality: Tracking theories versus phenomenal intentionality theories

Angela Mendelovici and David Bourget\*

Forthcoming in *Philosophy Compass*

Penultimate draft

*Abstract.* This paper compares and relates tracking and phenomenal theories of intentionality with respect to the issue of naturalism. Tracking theories explicitly aim to naturalize intentionality, while phenomenal intentionality theories do not. It might seem that considerations of naturalism count in favor of tracking theories. We survey key considerations relevant to this claim, including some motivations for and objections to the two kinds of theories. We conclude by suggesting that naturalistic considerations may in fact support phenomenal theories over tracking theories.

Some mental states are “of” or about something. For example, your belief that Pegasus has wings seems to be about Pegasus and his wingedness, and your visual experience of a cup presents or is about a cup. This directness of mental states is *intentionality*. *Intentional states* are states that exhibit intentionality; what they represent is their *content*.

The 1980s and 1990s saw the rise of *tracking theories* of intentionality, theories that aim to understand intentionality in terms of tracking or suitably co-varying with items in the environment. These theories have been motivated in good part by a perceived need to “naturalize” intentionality. Though tracking theories faced various challenges, many were considered to be merely technical, and research proceeded with much optimism. Tracking theories arguably remain the most popular theories of intentionality today.

The late 90s to the present saw the rise of a very different kind of theory of intentionality, *phenomenal intentionality theories*. These theories aim to understand intentionality in terms of another mental phenomenon, *phenomenal consciousness*, the “what it’s like” of mental states. Unlike tracking theories, which take intentionality to be a matter of how we are related to features of our environments, most phenomenal intentionality theories are radically internalistic. And unlike tracking theories, phenomenal intentionality theories are not primarily concerned with naturalizing intentionality.

It might seem that the recent interest in phenomenal intentionality theories represents a turn away from naturalism. Our aim in this paper is to survey the considerations relevant to this claim. Section 1 clarifies the notion of intentionality. Section 2 describes tracking theories and their goal of naturalizing intentionality. Section 3 describes objections to tracking theories. Section 4 describes phenomenal intentionality theories. Finally, section 5 overviews different conceptions of naturalism and suggests that considerations of naturalism may in fact favor phenomenal intentionality theories over tracking theories.

---

\* This paper is thoroughly co-authored.

# 1. Intentionality

A theory of intentionality aims to explain intentionality. Before we provide a theory of something, we need a way to fix on our target, on what it is that we are trying to provide a theory *of*.

There are various ways of fixing reference on intentionality. According to some, intentionality is a phenomenon that we notice in ourselves in the first instance (see e.g. Mendelovici 2010, Kriegel 2011). We notice that we have perceptual experiences of a blue cup on a table, or that we have judgments that Santa Claus is jolly. Our mental states introspectively seem to present, represent, or “say something” about the world, about how it is, how it might be, or how we would like it to be. While states that we cannot introspect upon might have intentionality, our reference-fixing grip on our target is through introspection.

Others, following Sellars (1956), take intentionality to be a theoretical posit of a folk theory of mind and behavior, *folk psychology* (see e.g. Fodor 1987). On this approach, beliefs, desires, and other kinds of intentional states are posits that help explain behavior. Many theorists who favor the folk psychological approach to fixing reference to intentionality also take intentionality to be a posit in various brands of cognitive science (see especially Fodor 1987), though it is unclear whether this alleged role in cognitive science is part of how they intend to fix reference on their target.<sup>1</sup>

We wish to remain neutral between the preceding approaches. The important point to observe is that both approaches allow that we have some pre-theoretic grip on at least some instances of intentionality, whether through introspection or folk intuition. Of course, this does not preclude cases of intentionality that we wouldn't introspectively or intuitively recognize as such.

## 2. Tracking theories of intentionality

Tracking theories of intentionality claim that intentionality is a tracking relation, where tracking is a matter of detecting, carrying information about, or otherwise corresponding with states or items in the environment (Stampe 1977, Dretske 1981, 1988, 1995, Millikan 1984, Papineau 1984, Fodor 1987, 1990, Neander 1996, Rupert 1999). Tracking theories have been applied to conceptual representation as well as to perceptual representation.<sup>2</sup>

---

<sup>1</sup> Though most theorists aiming to naturalize intentionality adopt a broadly intuitive approach to fixing reference on their target, Cummins (1994) is an exception in that he is interested in developing a notion of intentionality that is exclusively suitable for cognitive science.

<sup>2</sup> Several theorists have combined tracking theories of intentionality that apply to perceptual representation with intentionalism about phenomenal consciousness, the view that phenomenal consciousness is a species of intentionality. The resulting view is sometimes called *tracking representationalism*. See Lycan (1996), Dretske (1995), and Tye (2000) for defenses of this view. See also Bourget & Mendelovici (2014) for an overview.

On Fodor's (1987) view, a representation R represents content C just in case Cs cause Rs, and for any other Ds that cause Rs, the D-to-R connection is asymmetrically dependent on the C-to-R connection. Dependence is cashed out counterfactually: a causal connection A depends on another causal connection B just in case if B were to break, A would also break. For example, on Fodor's view, the representation HORSE represents *horse* and not *pony* because if the horse-to-HORSE connection were to break, then the pony-to-HORSE connection would also break, but if the pony-to-HORSE connection were to break, the horse-to-HORSE connection would remain intact. Intuitively, HORSE represents *horse* because horses are the most robust causes of HORSE.

On Millikan's (1984) view, tracking is not a causal relation, but a kind of correspondence relation. Simplifying a little, a representation represents whatever environmental condition the system that uses the representation (the representation's *consumer*) needs to be in place in order to perform its proper function. Proper functions are cashed out in evolutionary terms; a system's proper function is whatever it did in the system's ancestors that caused it to be selected for. Dretske (1988) also appeals to mental representations' functions. Simplifying again, his view is that representations represent what they have the function of indicating.<sup>3</sup>

One of the main virtues of tracking theories is that they provide a naturalistic account of intentionality. They tell us what intentionality *really is* using only naturalistically acceptable entities, such as causal relations, evolutionary histories, and the like. Indeed, the perceived need to naturalize intentionality is one of the main motivations for tracking theories. Fodor (1987) puts the point as follows:

I suppose that sooner or later the physicists will complete the catalogue they've been compiling of the ultimate and irreducible properties of things. When they do, the likes of *spin*, *charm*, and *charge*, will perhaps appear upon their list. But *aboutness* surely won't; intentionality simply doesn't go that deep. It's hard to see, in face of this consideration, how one can be a Realist about intentionality without also being, to some extent or other, a Reductionist. If the semantic and the intentional are real properties of things, it must be in virtue of their identity with (or maybe of their supervenience on?) properties that are themselves *neither* intentional *nor* semantic. If aboutness is real, it must really be something else. (p. 97, emphasis in original)

In other words, intentionality is real but not fundamental, so it must ultimately be reducible to some combination of fundamental physical ingredients. Millikan (1984, p. 87) and Dretske (1988, p. x) also present this kind of naturalization project as an important motivation for their tracking theories.

If naturalism is a key motivation for tracking theorists, then why do they identify intentionality with a tracking relation rather than some other plausibly physical property or relation? Seminal arguments made by Putnam (1975), Burge (1979), and Kripke (1980) have convinced many that words and concepts refer to external entities simply in virtue of appropriate causal connections they bear to these entities. From such a causal theory of reference, it is only a short step to tracking theories of intentionality. Fodor

---

<sup>3</sup> Dretske (1981) offers a different theory.

(1987, p. 98) and Millikan (1986) acknowledge the role of these arguments in motivating their approaches.

### 3. Objections to the tracking theory

Though tracking theories remain popular today, research on them has dwindled in the past two decades. Tracking theories face numerous objections, most of which can be understood as alleged counterexamples, cases in which tracking theories allegedly make the wrong predictions as to what a mental state represents.

The well-known disjunction problem can be understood in this way. The disjunction problem arises most clearly for what Fodor (1987) calls the *crude causal theory*, a theory that no one actually holds but that serves as a foil for various more plausible theories. The crude causal theory states that a representation represents whatever reliably causes it. Suppose that horses cause the representation we usually use when we think about horses, HORSE, but so too do cows on a dark night. But then, according to the crude causal theory, HORSE represents *horse or cow on a dark night*, since this is what reliably causes it. But this is the wrong answer; HORSE represents the non-disjunctive content *horse*. This inability to distinguish cases of misrepresentation from the representation of genuinely disjunctive contents is the *disjunction problem*.

The problem is not simply that the crude causal theory attributes contents that are disjunctive. Some representations might genuinely be disjunctive; for instance, someone might have a concept with the content *mouse or rat*. The problem, instead, is that the theory attributes disjunctive contents *when it should not*. Based on our pre-theoretic knowledge of intentionality, we know that the concept we use when we think about horses does not represent *horse or cow on a dark night*. Any theory that claims otherwise simply gets things wrong.

While no one endorses the crude causal theory, the disjunction problem also afflicts more plausible tracking theories. Dretske (1981) (but not the later Dretske) develops a theory on which representations represent whatever causes or would cause them during a specified learning period. Since cows on a dark night can cause HORSE in the learning period, the theory yields the false prediction that HORSE represents *horse or cow on a dark night* (Fodor 1987, pp. 102-4). Fodor (1990, p. 71) claims that Millikan's theory also suffers from the disjunction problem. Fodor's theory faces potential counterexamples with a similar flavor. One objection is that it has the unintended consequence that HORSE represents proximal sensory projections of horses (see Sterelny 1990, Antony & Levine 1991, Adams & Aizawa 1997, and Fodor 1990). Difficulties such as these might initially have seemed like superficial issues requiring at worst minor tweaks to the theory. However, they have proliferated and persisted. This history of failure suggests that the issues are not merely technical.

The preceding objections involved actual cases that are difficult for tracking theories to account for. Other potential counterexamples to tracking theories involve merely possible cases. Insofar as

tracking theories make predictions about such cases, these counterexamples are problematic. For instance, the merely possible case of swamp-persons is a potential counterexample to teleological tracking theories, theories on which evolutionarily-determined biological functions play a role in fixing content (e.g. Millikan's 1984 and Dretske's 1988). It seems possible in principle that a molecule-for-molecule duplicate of you could emerge from a swamp by pure chance (perhaps due to a random quantum event). On Millikan and Dretske's views, this *swamp-person* would not have any representational states, since it would lack an evolutionary history.<sup>4</sup> But this duplicate of you would enter exactly the same brain states and behave in exactly the same ways as you, at least initially. It seems absurd to think that it does not represent. Brown (1993), Baker (2007), and Pietroski (1992) develop this objection. Millikan (1996) bites the bullet on swamp-persons, claiming they don't have intentional states.

A variant of the swamp-person example extends to non-teleological tracking theories. Suppose that an isolated duplicate of your brain spontaneously appears in a life-sustaining box in space. This brain in a box has no potential for outside interaction. It intuitively seems it will initially have many of the same mental states as you, though your mental histories will quickly diverge, since it receives no outside stimulation, whereas you do. These intuitions might come from common sense or from minimal expert knowledge of the workings of the nervous system. The problem is that this brain in a box does not plausibly track anything, so tracking theories predict that it does not represent at all. This seems to be the wrong answer. See also Horgan, Tienson & Graham (2004), who argue against tracking theories on the basis of various intuitions about brains in vats.

Another family of objections to tracking theories argues that there is a general mismatch between what a representation tracks and what it represents. This type of objection works best with the contents of perceptual experiences. For example, color experiences seem to track objects' reflection of certain wavelengths of light, or their *reflectance properties*. But, as noted by Campbell (1969), Hardin (1988), and Maund (1995), there are important structural differences between reflectance properties and the contents of color experience. For example, purple as perceptually represented appears to be a binary hue, while blue is a unitary hue. Nothing in the structure of reflectance properties reflects this. This mismatch in structural properties suggests that the color properties we represent in perceptual experience are not reflectance properties, hence not things that we track (assuming that what we track are reflectance properties).

One variant on this type of argument emphasizes differences in the similarity relations that hold between represented contents and tracked properties. Pautz (2013b) presents an argument of this kind, though with a different target in mind.<sup>5</sup> Pautz points to such examples as aspartame and its stereoisomers. Aspartame and some of its stereoisomers are chemically very similar but taste very different. The contents we represent when tasting them are very different while the properties tracked are very similar. Conversely, some taste experiences track very different chemical compounds but represent similar

---

<sup>4</sup> This thought experiment was first introduced by Davidson (1987), but for a different purpose.

<sup>5</sup> Pautz' argument targets tracking representationalism, a view that combines a tracking theory of intentionality with an account of the phenomenal character of experiences in terms of intentionality (see also fn. 2). However, a similar argument can be run against tracking theories of intentionality. We present this argument in the main text.

contents. For instance, a wide variety of chemical compounds cause bitter tastes. Such mismatches in similarity relations between tracked contents and experienced contents suggest that perceptual representation does not require tracking.

Another kind of mismatch case involves a mismatch in the internal structure and quality of tracked properties and represented contents. Mendelovici (2010) argues that there are many paradigm cases of representations that do not behave as if they represent what they track. For example, perceptual representations of color might plausibly be said to track dispositions to reflect certain proportions of certain wavelengths of light, dispositions to cause certain psychological states in us, or the physical categorical bases of the aforementioned dispositions. The properties tracked are complex, sophisticated, and sometimes dispositional or relational properties. But our color representations do not in any way behave as if they represented such properties. First, such potential content attributions don't capture how things seem to us when they seem colored. Second, our perceptual representations do not make contents involving such complex properties available to belief-forming mechanisms or other cognitive processes. For instance, from perceptually experiencing something as blue<sup>421</sup>, we are not able to infer that it has any of the properties it can be said to track. Lastly, our behavior towards objects does not in general betray a representation of them as having the relevant complex, sophisticated, relational, or dispositional properties. If all this is right, then our perceptual representations of colors simply does not behave as if it represent the properties they track, and so, arguably, they do not represent what they track. Other potential mismatch cases include conceptual representations of color, perceptual representations of hotness and coldness, and representations of moral properties. As in the case of the disjunction problem and swamp-person, mismatch cases are cases in which tracking theories make the wrong predictions.

Some tracking theorists take tracked properties to be represented under certain modes of presentation (e.g. Fodor 1987) or, at least in some perceptual cases, non-conceptually (Tye 2000). Perhaps, then, modes of presentation or non-conceptual representation transform tracked contents into the contents we intuitively or introspectively grasp. But it is far from clear that modes of presentation or non-conceptual representation can help here. The problem with modes of presentation is that they now seem to do all the work in accounting for content as we understand it, which threatens to make tracked contents explanatorily idle. The problem with non-conceptual content is that it is just not clear how representing something without needing to have a concept for it can transform utterly foreign contents into contents that are introspectively and intuitively familiar.<sup>6</sup>

The tracking theory aims to account for intentionality in terms of relations to the environment. But, as we have seen, it is difficult to specify tracking relations that are able to connect us to the right things. In particular, tracking theories do not seem to be able to connect us to what seem to be the contents of our mental states from a pre-theoretic point of view.<sup>7</sup> This points towards a different approach to intentionality, one that locates the source of intentionality not in our connections to the environment, but in our own subjective experiences. The next section explores such an approach.

---

<sup>6</sup> See also Mendelovici (2013), who argues that merely psychologically possible mismatch cases are also problematic for tracking theories.

<sup>7</sup> Similarly, Mark Johnston (2007) argues that tracking theories cannot explain the contents we are manifestly presented with.

## 4. Phenomenal intentionality theories

The 1990s saw the rise of a kind of theory radically different from the tracking theory, the *phenomenal intentionality theory* (PIT), on which intentionality is somehow or other a matter of phenomenal consciousness. Proponents of PIT (or something close to it) include Loar (1987, 1988, 1995, 2002, 2003), Searle (1990, 1991, 1992), Goldman (1993), Strawson (1994), Siewert (1998), McGinn (1988), Horgan & Tienson (2002), Horgan, Tienson & Graham (2003, 2004, 2006), Georgalis (2006), Pitt (2004, 2009, 2011), Farkas (2008a, 2008b, 2013), Bourget (2010), Mendelovici (2010), Kriegel (2007, 2011, 2013), Horgan & Graham (2012), Horgan (2013), and Pautz (2013a).

*Phenomenal consciousness* is the subjective or qualitative aspect of mental states, or the “what it’s like” of being in certain mental states (see Nagel, 1974). Mental states exhibiting phenomenal consciousness are *phenomenal states*, and can be said to have *phenomenal properties*. Phenomenal properties are characterized by their *phenomenal character*, their particular “what it’s like.” Examples of phenomenal states include conscious visual experiences, pains, emotional feelings, and cognitive experiences such as the feeling of *déjà vu*.

Advocates of PIT claim that there are *phenomenal intentional states*, intentional states that are either identical to or directly grounded in phenomenal states.<sup>8</sup> *Phenomenal intentionality* is a kind of intentionality that arises from consciousness in this way. The most plausible examples of phenomenal intentional states are perceptual. For example, once you have a state with a certain bluish phenomenal character (a certain phenomenal property), it seems that you thereby represent a particular shade of blue. We call the thesis that there is phenomenal intentionality *weak PIT*.

(Weak PIT)     There is phenomenal intentionality.

Proponents of weak PIT usually take consciousness to be in some sense prior to phenomenal intentionality, though some versions of the view that endorse an identity between phenomenal intentional states and phenomenal states take the two to be on par (Pautz 2008, Bourget 2010, Mendelovici 2010).

There is disagreement regarding the extent of phenomenal intentionality. However, many proponents of weak PIT agree that perceptual states and occurrent cognitive states (such as thoughts, beliefs, and desires that one is currently entertaining) have phenomenal intentionality, and that non-conscious, subpersonal, and non-occurrent states, such as beliefs that one is not currently entertaining, do not (see e.g. Pitt 2008, Horgan & Tienson 2002).

---

<sup>8</sup> By “states” we mean token states. This understanding of phenomenal intentional states allows that some tokens of certain types of intentional states are phenomenal intentional while others are not. For example, it allows that one can represent redness both through a phenomenal intentional perceptual state and through a non-phenomenal cognitive state.

Phenomenal intentionality theorists generally endorse the further claim that all non-phenomenal forms of intentionality, if there are any, are in some way at least partially derived from phenomenal intentionality. We call this stronger claim *strong PIT*. A more extreme view denies that there are non-phenomenal forms of intentionality. We call this view *extreme PIT*.

(Strong PIT) There is phenomenal intentionality, and all non-phenomenal intentionality derives (at least in part) from phenomenal intentionality.

(Extreme PIT) There is phenomenal intentionality, and all intentionality is phenomenal intentionality.

Most phenomenal intentionality theorists endorse strong PIT but not extreme PIT. Amongst proponents of strong PIT, there is disagreement on the extent of non-phenomenal intentionality. Some claim that non-conscious or subpersonal states, such as states in early visual processing, exhibit non-phenomenal intentionality that derives from their connections to phenomenal intentional states (see e.g. Horgan and Tienson 2002, and Kriegel 2001), whereas others deny that such states are intentional at all (Strawson 1994, 2011a, and Mendelovici 2010). Loar (2003), Bourget (2010), and Mendelovici (2010) take at least some of the content of occurrent cognitive states, such as thoughts, to be derived from their relations to other states, whereas most other advocates of PIT take the intentional properties of such states to be fully determined by their phenomenal properties (see e.g. Pitt 2004).

The guiding idea behind the phenomenal intentionality research program is that consciousness is the source of all intentionality, so, unless otherwise noted, we will take “PIT” to refer to strong PIT, the weakest view that captures this idea.

One might ask whether it is appropriate to regard PIT as an *alternative* to tracking theories. One reason to think the two theories are not alternatives is that they are in fact compatible. For example, it could turn out that phenomenal consciousness is the source of all intentionality and phenomenal consciousness reduces to tracking (Kriegel 2011 defends a view of this type).

The choice between PIT and tracking theories might not be an exclusive one, but this does not mean that we should not think of them as alternatives. They are alternative theories in the same sense as the theory that the Golden Gate Bridge is made of steel (STEEL) and the theory that it is made of gold (GOLD) are alternative theories. For all we know, it could turn out that gold is a kind of steel,<sup>9</sup> so the choice between GOLD and STEEL is not (logically) mutually exclusive. This does not mean that the theories are not distinct, as different considerations weigh for and against them (the Linguistic Argument for GOLD, the Engineering Argument for STEEL). Similarly, phenomenal consciousness *might* reduce to

---

<sup>9</sup> Note that this is compatible with it being an a posteriori necessary truth that gold is not steel. All we are saying is that this truth is a posteriori.



tracking, but it might not (at the moment, a majority of theorists do not believe that it does).<sup>10</sup> Different considerations weigh for and against these theories, so they should be treated as distinct.

Another reason one might think that PIT is not an alternative to tracking theories is that the two theories do not aim to do the same job. One might suggest that PIT aims to explain the relationship between consciousness and intentionality, while tracking theories aim to provide a naturalistic account of intentionality. It's true that PIT and tracking theories can serve different purposes (this is true merely in virtue of the fact that they say different things), but they do share the overarching aim of providing a theory of intentionality: they both aim to tell us what intentionality really *is*.<sup>11</sup>

PIT avoids some of the worries with the tracking theory described in the previous section. If we assume that phenomenal properties are intrinsic properties in that they are determined by what goes on in subjects from the skin-in, then swamp-person and other molecule-for-molecule duplicates of you have the same phenomenal intentional properties as you. On versions of PIT that accept derived intentionality, these duplicates might also share your non-phenomenal intentional properties, depending on how exactly non-phenomenal intentionality is derived from phenomenal intentionality.<sup>12</sup>

For similar reasons, PIT avoids the mismatch problem. What representations represent is just what we experience when we token them. In the case of colors, this might be primitive or *sui generis* color properties, or what Chalmers (2006) calls "Edenic colors," colors as they naively appear to be. This seems to match what we perceptually represent when we represent colors. An advocate of PIT need not accept realism about Edenic colors; she can claim that color experiences systematically misrepresent.<sup>13</sup> Something similar can be said about other mismatch cases.<sup>14</sup>

---

10 The main view on which consciousness reduces to a kind of tracking is tracking representationalism (see e.g. Dretske 1995, Tye 2000). According to the survey performed by Bourget & Chalmers (forthcoming), representationalism has an acceptance rate of only 31% among professional philosophers. While the acceptance rate of tracking representationalism was not measured, we should expect it to be significantly lower than 31%, because many representationalists are not *tracking* representationalists (see Bourget & Mendelovici, 2013).

11 Note that neither kind of theory does this by accounting for intentionality in terms of fundamental entities (for example, basic particles). PIT does not tell us how consciousness relates to fundamental entities, and tracking theories do not account for their tracking relations in terms of fundamental entities. For instance, Fodor's theory appeals to facts about counterfactuals, which are not themselves explained and seem difficult to naturalize.

12 Horgan and Tienson (2002), Bourget (2010), and Mendelovici (2010) allow that environmental factors can make a difference to what some states derivatively represent. These versions of PIT can accommodate some externalist intuitions. For example, they can allow that intrinsic duplicates can express different thoughts with "Water is wet."

13 If the advocate of PIT endorses a tracking theory of consciousness, she will have trouble being an anti-realist about Edenic colors for the same reason that tracking theories of intentionality have trouble maintaining color anti-realism (see Mendelovici 2013).

14 An anonymous reviewer has suggested that this response to the mismatch problem is also available to the tracking theorist. She need only accept a kind of primitivist color realism of the sort defended by John Campbell (1993). The difference between PIT's color primitivism and this tracking theorist's color primitivism is that PIT allows that primitive color properties are not instantiated, whereas the tracking theorist cannot easily allow for that (see Mendelovici 2013).

But PIT faces its own challenges. In order to be plausible, it has to provide a satisfactory treatment of the kinds of states that we might pre-theoretically consider to be intentional. This might involve arguing that they are intentional, arguing that they exhibit phenomenal intentionality, or showing how their intentionality derives from the phenomenal intentionality of other states. PIT also faces challenges in explaining how phenomenal intentional states can correctly or incorrectly connect with or match the external world, yielding reference, truth, and accuracy. Bourget & Mendelovici (forthcoming) survey these challenges and the various options for addressing them.

There is a more glaring worry with the entire project that is likely to preoccupy anyone at all attracted to tracking theories: PIT doesn't appear to be naturalistic. Though it does attempt to provide an account of intentionality in terms of *something else*, this something else is phenomenal consciousness, which, as Chalmers (1995, 1996), Levine (1983), and others have argued, is difficult to fit into a physicalist picture of the world. So, it might seem that PIT fails to naturalize intentionality, making it less attractive as a theory.

## 5. Naturalism about intentionality

So far, we've described tracking theories and PITs, and what we take to be some central virtues and vices of each. Tracking theories are naturalistic, but have trouble dealing with certain cases, while PITs can deal with the relevant cases, but appear not to be naturalistic. In this section, we want to suggest that this conflict is not between naturalism and avoiding counterexamples, but between different kinds of naturalistic considerations. We also suggest that, all things considered, naturalistic considerations in fact recommend PIT over the tracking theory.

Naturalism is a commitment motivated by the success of scientific inquiry. As Shapiro (1997), Chomsky (1995), Horst (2009), and Papineau (2007) have argued, there are different ways of interpreting this commitment. It can be interpreted as a commitment to a certain ontology delivered by science; this is *ontological naturalism*. But it can also be interpreted as a methodological commitment to a broadly scientific methodology; this is *methodological naturalism*.

Ontological naturalism presupposes a particular doctrine about what there is. There are mindless fundamental physical entities (for example, superstrings or fields), and combinations or arrangements of fundamental physical entities; call all such items *naturalistic items*. Ontological naturalism is a commitment to an ontology of naturalistic items. Ontologically naturalistic theories aim to account for various "problematic" items, such as moral properties, abstract entities, consciousness, and meanings in terms of naturalistic items.

Shapiro (1997) argues that the project of providing an ontologically naturalistic account of intentionality begs the question about what items are to be considered naturalistic. Chomsky (1995) argues that science does not even deliver a clear ontology in the first place, so there is no available list of naturalistic items in terms of which we should explain problematic items like intentionality. While we

think there is something right about these criticisms, they might strike some as overly skeptical or dismissive of naturalistic reductive projects in general.

We want to suggest that either PIT is ontologically naturalistic, or ontological naturalism is not an appropriate commitment for reductive projects. Here is why: Either consciousness reduces to naturalistic items or it does not. If consciousness reduces to naturalistic items, then PIT is ontologically naturalistic, since it appeals only to consciousness and other naturalistic items. Suppose, now, that consciousness does not reduce to naturalistic items. Since eliminativism about consciousness is not an option, we should accept consciousness as a new fundamental entity in our ontology. But then ontological naturalism does not provide the correct ontology; it is not true that the correct ontology includes only mindless physical entities and entities made up out of them. But then being ontologically naturalistic is not more virtuous than invoking an ontology of naturalistic entities plus consciousness.<sup>15</sup> Thus, if consciousness is not a naturalistic item, ontological naturalism is not an appropriate commitment. So, either PIT is ontologically naturalistic or ontological naturalism is not an appropriate commitment. Either way, ontological naturalism is not grounds for rejecting PIT.<sup>16</sup>

In a similar spirit, Strawson (2011b) argues that "real" naturalism should be realist about consciousness. While Strawson's naturalism is an ontological thesis, he rejects our definition of ontological naturalism as involving a commitment to an ontology of mindless physical entities and combinations of such entities, instead taking naturalism to be an ontological thesis that must make room for consciousness. On Strawson's conception of naturalism, PIT is wholly naturalistic.

Let us turn to methodological naturalism and its demands. Methodological naturalism is a commitment to a certain kind of broadly scientific methodology. It involves taking a broadly scientific perspective on inquiry. Some proponents of methodological naturalism are skeptical that methodological naturalism can be put in the form of a claim or set of claims. For example, Maddy writes:

[N]aturalism, as I understand it, is not a doctrine, but an approach; not a set of answers, but a way of addressing questions. As such, it can hardly be described in a list of theses: it can only be seen in action! (Maddy, 2001, p. 37)

Maddy adopts a broad conception of what the relevant scientific methodology involves, denying that we can characterize it in terms of necessary and sufficient conditions (Maddy, 2001, p. 48). Timothy

---

<sup>15</sup> What is virtuous about ontological naturalism (if anything) is not sticking to a naturalistic ontology per se, but rather sticking to an ontology of items that exist. If ontological naturalism provides the correct ontology, then our theories should be constrained to invoking only naturalistic items. But if we expand our ontology to include other items, then our theories should only be constrained to invoking items from this revised list.

<sup>16</sup> Even if ontological naturalism can't be grounds for rejecting PIT, it might be that PIT is grounds for rejecting ontological naturalism. Pautz (2010) argues that PIT leaves us no choice but to endorse primitivism about intentionality. This leads to the first branch of our dilemma: if consciousness is not reducible to the naturalistic, then PIT is incompatible with ontological naturalism about intentionality, but we should not count this against PIT. The reason this should not count against PIT is that ontological naturalism is a virtue only if consciousness is reducible to the naturalistic.

Williamson (2011) can also be understood as a methodological naturalist (though he would deny the title). He recommends what he calls “the scientific spirit,” which is open to various modes of inquiry depending on the subject matter.

There are many questions concerning the appropriate methodology for investigating intentionality. For example, there are questions concerning the relevance of thought experiments like the swamp-person thought experiment. But however exactly we understand methodological naturalism, it is clear that it at least involves a commitment to accommodating any available relevant evidence. Arguably, it also involves a commitment to flowing from that evidence, to following the evidence where it leads.

This is one way in which tracking theories seem to fail to be methodologically naturalistic; they fail to account for all the evidence. In section 2, we overviewed reasons for thinking that there are counterexamples to tracking theories. If these arguments are sound, then tracking theories do not account for all the data. They are not empirically adequate. This is clearest in the case of the mismatch problem. It’s unclear what kinds of contortions could avoid the apparently inaccurate predictions of tracking theories in mismatch cases. Since tracking theories do not account for all the evidence concerning intentionality, they fail to be methodologically naturalistic. PIT, in contrast, adequately deals with these cases and avoids this barrier to methodological naturalism.

Of course, PIT might have its own set of problem cases. In the previous section, we listed some challenging cases for PIT, such as cases of standing states and non-conscious or subpersonal states. Some versions of PIT claim that such states have non-phenomenal intentionality, which derives from their connections to phenomenally intentional states. Other versions deny that these states have intentionality at all. While the issues arising from such cases are far from being settled, we want to suggest that the status of these potentially problematic cases for PIT is different from the status of mismatch cases for tracking theories. It is doubtful that tracking theories can at all accommodate mismatch cases without denying the evidence, that is, denying that the relevant states represent what they manifestly represent. In the case of PIT and its challenging cases, it is not pre-theoretically obvious that the relevant states have *phenomenal* intentionality, rather than derived intentionality or, in some cases, no intentionality at all, and so we have no strong reason to think that PIT cannot handle them.

Following a suggestion by Kriegel (2011, pp. 172) about an analogous case,<sup>17</sup> one might construe the choice between tracking theories and PITs as boiling down to one between theoretical virtues and empirical adequacy, where the theoretical virtue in question is that of being ontologically naturalistic. Kriegel claims to feel the pull of both virtues, and ends up splitting his credence between the two approaches he is concerned with roughly equally. Earlier in this section, we argued that it is not clear that considerations of ontological naturalism support tracking theories over PIT. If that is right, then the only considerations left to factor in are those of empirical adequacy, which, we claim, support PIT. But even if we assume that tracking theories come out ahead when it comes to theoretical virtues, we submit that there is no real choice to be made between theoretical virtues and empirical adequacy. This is because,

---

<sup>17</sup> Kriegel is concerned with choosing between a higher-order tracking theory and an adverbial theory of phenomenal intentionality, but his points map onto our choice between tracking theories and PITs.

assuming our evidence is reliable, empirical inadequacy is fatal to a theory; it means that the theory as stated is false. Exemplifying various theoretical virtues cannot “make up” for being false. While ontological naturalism may be attractive, it doesn’t make up for a failure of empirical inadequacy.

## 6. Conclusion

In this paper, we presented tracking theories and phenomenal intentionality theories. We outlined what we consider to be their main virtues and their main problems. We then focused on the issue of naturalism and briefly argued that, perhaps surprisingly, considerations of naturalism favor PIT over tracking theories. Our discussion is cursory and much more needs to be said in order to establish the success or failure of any theory of intentionality. But we at least hope to have convinced the reader that considerations of naturalism do not cut against the phenomenal intentionality theory.<sup>18</sup>

## Bibliography

- Adams, Frederick R. & Aizawa, Kenneth (1997). Fodor's asymmetric causal dependency theory and proximal projections. *Southern Journal of Philosophy* 35 (4):433-437.
- Antony, Louise M. & Levine, Joseph (1991). The nomic and the robust. In Barry M. Loewer & Georges Rey (eds.), *Meaning in Mind: Fodor and His Critics*. Blackwell.
- Baker, Lynne Rudder (2007). Naturalism and the first-person perspective. In Georg Gasser (ed.), *How Successful is Naturalism? Publications of the Austrian Ludwig Wittgenstein Society*. Ontos Verlag.
- Bourget, David (2010). Consciousness is Underived Intentionality. *Noûs* 44 (1):32-58.
- Bourget, David & Mendelovici, Angela (2014). Tracking Representationalism. In Andrew Bailey (ed.), *Philosophy of Mind: The Key Thinkers*. Continuum.
- Bourget, David & Mendelovici, Angela (forthcoming). Phenomenal Intentionality, *Stanford Encyclopedia of Philosophy*.
- Bourget, David & Chalmers, David J. (forthcoming). What do philosophers believe? *Philosophical Studies*.
- Brown, Deborah J. (1993). Swampman of la mancha. *Canadian Journal of Philosophy* 23 (3):327-48.
- Burge, Tyler (1979). Individualism and the mental. *Midwest Studies in Philosophy* 4 (1):73-122.
- Campbell, John (1993). A simple view of colour. In John J. Haldane & C. Wright (eds.), *Reality: Representation and Projection*. OUP.
- Campbell, Keith (1969). Colours, in *Contemporary Philosophy in Australia*, ed. R. Brown and C. D. Rollins, London: Allen and Unwin: 132 – 57.
- Chalmers, David J. (2006). Perception and the fall from Eden. In Tamar S. Gendler & John Hawthorne (eds.), *Perceptual Experience*. Oxford University Press.
- Chalmers, David J. (2004). The representational character of experience. In Brian Leiter (ed.), *The Future for Philosophy*. Oxford University Press.

---

<sup>18</sup> Thanks to Adam Pautz and an anonymous referee for helpful comments.

- Chalmers, David J. (1995). Facing up to the problem of consciousness. 2 (3):200-19.
- Chomsky, Noam (1995). *The Minimalist Program*. The MIT Press.
- Cummins, Robert (1994). Interpretational semantics. In Stich, S. P. and Warfield, T. A., editors, *Mental Representation: A Reader*, pages 278–301. Blackwell, Oxford.
- Davidson, Donald (1987). “Knowing One’s Own Mind”, *Proceedings and Addresses of the American Philosophical Association* 60, 441-458.
- Dretske, Fred (1995). *Naturalizing the Mind*. MIT Press.
- Dretske, Fred (1988). *Explaining Behavior: Reasons in a World of Causes*. MIT Press.
- Dretske, Fred (1981). *Knowledge and the Flow of Information*. MIT Press.
- Farkas, Katalin (2013). Constructing a World for the Senses. In Uriah Kriegel (ed.), *Phenomenal Intentionality*. OUP.
- Farkas, Katalin (2008a). Phenomenal intentionality without compromise. *The Monist* 91 (2):273-93.
- Farkas, Katalin (2008b). *The Subject's Point of View*. Oxford University Press.
- Fodor, Jerry A. (1990). *A Theory of Content and Other Essays*. MIT Press.
- Fodor, Jerry A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press.
- Georgalis, Nicholas (2006). Representation and the first-person perspective. *Synthese* 150 (2):281-325.
- Goldman, Alvin (1993). The psychology of folk psychology. *Behavioral and Brain Sciences* 16:15-28.
- Horgan, Terence (2013). Original Intentionality is Phenomenal Intentionality. *The Monist* 96 (2):232-251.
- Horgan, Terence & Graham, George (2012). Phenomenal Intentionality and Content Determinacy. In Richard Schantz (ed.), *Prospects for Meaning*. De Gruyter.
- Johnston, Mark (2007). Objective mind and the objectivity of our minds. *Philosophy and Phenomenological Research* 75 (2):233–268.
- Hardin, C. L. (1988). *Color for Philosophers: Unweaving the Rainbow*, Indianapolis: Hackett.
- Horst, Steven (2009). Naturalisms in philosophy of mind. *Philosophy Compass* 4 (1):219-254.
- Horgan, Terence E. & Tienson, John L. (2002). The intentionality of phenomenology and the phenomenology of intentionality. In David J. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press.
- Horgan, Terence E., Tienson, John L. & Graham, George (2006). Internal-world skepticism and mental self-presentation. In Uriah Kriegel & Kenneth Williford (eds.), *Self-Representational Approaches to Consciousness*. MIT Press.
- Horgan, Terence E., Tienson, John L. & Graham, George (2004). Phenomenal intentionality and the brain in a vat. In Richard Schantz (ed.), *The Externalist Challenge*. Walter De Gruyter.
- Horgan, Terence E., Tienson, John L. & Graham, George (2003). The phenomenology of first-person agency. In Sven Walter & Heinz-Dieter Heckmann (eds.), *Physicalism and Mental Causation*. Imprint Academic.
- Kriegel, Uriah (2013). The Phenomenal Intentionality Research Program. In U. Kriegel (ed.), *Phenomenal Intentionality*. Oxford University Press.
- Kriegel, Uriah (2011). *The Sources of Intentionality*. Oxford University Press.
- Kriegel, Uriah (2007). Intentional inexistence and phenomenal intentionality. *Philosophical Perspectives* 21 (1):307-340.
- Kripke, Saul A. (1980). *Naming and Necessity*. Harvard University Press.

- Levine, Joseph (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly* 64 (October):354-61.
- Loar, Brian (2003). Phenomenal intentionality as the basis of mental content. In Martin Hahn & B. Ramberg (eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge*. MIT Press.
- Loar, Brian (2003). Transparent experience and the availability of qualia. In Quentin Smith & Aleksandar Jokic (eds.), *Consciousness: New Philosophical Perspectives*. Oxford University Press.
- Loar, Brian (1995). Reference from the first person perspective. *Philosophical Issues* 6:53-72.
- Loar, Brian (1988). Social content and psychological content. In Robert H. Grimm & D. D. Merrill (eds.), *Contents of Thought*. University of Arizona Press.
- Loar, Brian (1987). Subjective intentionality. *Philosophical Topics* 15 (1):89-124.
- McGinn, Colin (1988). Consciousness and content. *Proceedings of the British Academy* 74:219-39.
- Maund, Barry (1995). *Colours: Their Nature and Representation*, New York: Cambridge University Press.
- Mendelovici, Angela (2013). Reliable misrepresentation and tracking theories of mental representation. *Philosophical Studies* 165 (2):421-443.
- Mendelovici, Angela (2010). *Mental Representation and Closely Conflated Topics*. Dissertation, Princeton University.
- Millikan, Ruth G. (1996). On swampkinds. *Mind and Language* 11 (1):103-17.
- Millikan, Ruth G. (1986). Thoughts without laws: Cognitive science with content. *Philosophical Review* 95 (January):47-80.
- Millikan, Ruth G. (1984). *Language, Thought and Other Biological Categories*. MIT Press.
- Montague, Michelle (2010). Recent Work on Intentionality. *Analysis* 70 (4):765-782.
- Nagel, Thomas (1974). What is it like to be a bat? *Philosophical Review* 83 (October):435-50.
- Neander, Karen (1996). Swampman meets swampcow. *Mind and Language* 11 (1):118-29.
- Papineau, David (2010). Naturalism. *Stanford Encyclopedia of Philosophy*.
- Papineau, David (1984). Representation and explanation. *Philosophy of Science* 51 (December):550-72.
- Pautz, Adam (2006). Can the physicalist explain colour structure in terms of colour experience? *Australasian Journal of Philosophy* 84 (4):535 – 564.
- Pautz, Adam (2008). The Interdependence of Phenomenology and Intentionality. *The Monist* 91 (2):250-272.
- Pautz, Adam (2010). A Simple View of Consciousness, in G. Bealer and R. Koons (eds.), *The Waning of Materialism*. Oxford University Press.
- Pautz, Adam (2013a). Is phenomenology the ground of intentionality? In Uriah Kriegel (ed.), *Phenomenal Intentionality*. Oxford.
- Pautz, Adam (2013b). The real trouble for phenomenal externalists: New empirical evidence for a brain-based theory of consciousness. In Richard Brown (ed.), *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*. pp. 237-298.
- Pitt, David (2011). Introspection, Phenomenality, and the Availability of Intentional Content. In Tim Bayne & Michelle Montague (eds.), *Cognitive Phenomenology*. OUP.
- Pitt, David (2009). Intentional psychologism. *Philosophical Studies* 146 (1):117 – 138.
- Pitt, David (2004). The phenomenology of cognition, or, what is it like to think that P? *Philosophy and Phenomenological Research* 69 (1):1-36.

- Putnam, Hilary (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science* 7:131-193.
- Pietroski, Paul (1992). "Intentionality and Teleological Error," *Pacific Philosophical Quarterly* 73 (3): 267-82.
- Rupert, Robert D.(1999). The best test theory of extension: First principle(s). *Mind and Language* 14 (3):321-355.
- Searle, John R. (1994). The connection principle and the ontology of the unconscious: A reply to Fodor and Lepore. *Philosophy and Phenomenological Research* 54 (4):847-55.
- Searle, John R. (1992). *The Rediscovery of the Mind*. MIT Press.
- Searle, John R. (1991). Consciousness, unconsciousness and intentionality *Philosophical Issues* 1 (1):45-66.
- Searle, John R. (1990). Consciousness, explanatory inversion and cognitive science. *Behavioral and Brain Sciences* 13:585-642.
- Searle, John R. (1987). Indeterminacy, empiricism, and the first person. *Journal of Philosophy* 84 (3):123-146.
- Sellars, Wilfrid S. (1956). Empiricism and the philosophy of mind. *Minnesota Studies in the Philosophy of Science* 1:253-329.
- Shapiro, Lawrence A. (1997). The nature of nature: Rethinking naturalistic theories of intentionality. *Philosophical Psychology* 10 (3):309-322.
- Siewert, Charles (1998). *The Significance of Consciousness*. Princeton University Press.
- Stampe, Dennis W. (1977). Towards a causal theory of linguistic representation. *Midwest Studies in Philosophy* 2 (1):42-63.
- Sterelny, Kim (1990). *The Representational Theory of Mind*. Blackwell.
- Strawson, Galen (2011a). Cognitive phenomenology: real life. In Tim Bayne & Michelle Montague (eds.), *Cognitive phenomenology*. Oxford University Press.
- Strawson, Galen (2011b). Real Naturalism. In *Proceedings and Addresses of the APA* 86:2
- Strawson, Galen (1994). *Mental Reality*. MIT Press.
- Tye, Michael (2000). *Consciousness, Color, and Content*. MIT Press.
- Williamson, Timothy (2011). "What is Naturalism?" *The New York Times (The Stone)*, September 4. Retrieved online at <<http://opinionator.blogs.nytimes.com/2011/09/04/what-is-naturalism/>>