**Nonnatural Mental Representation[[1]](#footnote-1)**

Gualtiero Piccinini

To appear in J. Smortchkova, K. Dolega, T. Schlicht, eds., *What Are Mental Representations?* Oxford: Oxford University Press.

**Abstract**. I distinguish between two types of representation, natural and nonnatural. I argue that nonnatural representation is necessary to explain intentionality. I also argue that traditional accounts of the semantic content of mental representations are insufficient to explain nonnatural representation and, therefore, intentionality. To remedy this, I sketch an account of nonnatural representation in terms of natural representation plus offline simulation of nonactual environments plus tracking the ways in which a simulation departs from the actual environment. This is a step towards a naturalistic, mechanistic, neurocomputational account of intentionality.

**1. Intentionality**

Here are two things that minds can do. First, they can think about objects that may or may not exist; second, they can think about both truths and falsehoods. These are not the same capacity: the truths and falsehoods that minds can think are about things that either exist or don’t exist. Compare: “I exist,” “I don’t exist,” “there is a unicorn,” “there are no unicorns.”

Some minds are capable of using a natural language. In addition to thinking, linguistically endowed minds can talk—either truly or falsely—about things that may or may not exist. And talking truly or falsely is not the only kind of talking. In addition to asserting, as in “there is a unicorn,” utterances may have other kinds of *force*: inquiring, ordering, etc. So, utterances have both meaning—semantic content—and force.

Thinking and talking are aspects of intentionality, which is notoriously difficult to explain. Philosophers have tried for a long time and made some progress, but they have yet to find a complete explanation. Rather than reviewing the debate, I will begin with what I take to be the most promising explanatory strategy—the representational theory of intentionality—and propose a way to improve over previous versions.[[2]](#footnote-2)

In the next section, I introduce the representational theory of intentionality. In Section 3, I distinguish two kinds of (descriptive) representation: natural and nonnatural. I argue that nonnatural representation is necessary to fully explain intentionality. In Section 4, I briefly introduce my preferred mechanistic, neurocomputational theoretical framework—including the notion of structural representation and the functional role of structural representations, which is to simulate their target in order to guide action. In Section 5, I introduce a version of Informational Teleosemantics, which I take to be the most adequate theory of the content of (descriptive) mental representation. In Section 6, I argue that Informational Teleosemantics fails to explain nonnatural representation and hence to fully explain intentionality. In Section 7, I introduce offline simulation, which will help build an account of nonnatural representation. In Section 8, I sketch an account of the semantic content of offline simulations. Finally, in Section 9 I extend traditional accounts of mental representation by sketching an account of nonnatural representation in terms of offline simulation. This is a step towards a naturalistic, mechanistic, neurocomputational account of intentionality.

**2. The Representational Theory of Intentionality**

Explaining intentionality requires explaining both how linguistic utterances (and other signs) gain their meaning and how mental states such as beliefs and desires gain their semantic content. The Mentalist Theory of Meaning (MTM) maintains that the meaning of utterances can be explained in terms of the semantic content of mental states (cf. Speaks 2017, Section 3). If this is right, then in order to explain intentionality all we have to do is explain the semantic content of mental states.

More precisely, MTM claims that utterances gain both their force and their meaning in virtue of speakers possessing mental states with the relevant attitudes and semantic contents:

Utterance U has force F and means that P as a function of the speaker possessing mental state M with attitude A and content P.

There are several versions of MTM, which differ on how the meaning and force of utterances depends on the content and attitude of appropriate mental states. Here I cannot review the options but I can illustrate with a couple of examples. An utterance of “c’è un unicorno” is an *assertion* that there is a unicorn because it expresses the speaker’s belief that there is a unicorn. The attitude expressed by the assertion is *belief*, and the belief’s (and hence the assertion’s) content is *that there is a unicorn*. By contrast, an utterance of “mi regali un unicorno?” is a *request* that the hearer gift a unicorn to the speaker because it expresses the speaker’s desire that the hearer gift her a unicorn. Here, desire is the attitude expressed by the request, and the desire (and hence the request’s) content is *that the hearer gifts a unicorn to the speaker*.

Assuming that some version of MTM is correct, what remains to be explained is the attitudes and semantic content of mental states. The Representational Theory of Intentionality (RTI) claims that the attitudes and semantic contents of mental states are explained in terms of internal states, called representations, with appropriate functional roles and semantic contents.[[3]](#footnote-3) The explanation takes roughly the following form:

Mind M has mental state with attitude A and content P if and only if it possesses an internal representation with functional role RA and content P.

Here, RA is the functional role that corresponds to attitude type A. For example, beliefs are explained by mental representations with an appropriate functional role RBel, such that, among other effects, under appropriate circumstances they cause the utterance of assertions with the same content that the belief has. By contrast, desires are explained by mental representations with a different functional role RDes, such that, among other effects, under appropriate circumstances they cause the utterance of requests with the same content that the desire has. In addition to beliefs and desires, RTI may posit other types of mental representations corresponding to other mental states.

RTI is behind much work in the philosophy of mind and language of recent decades (Jacob 2014, Speaks 2017, Pitt 2017). RTI also fits well with the way mainstream cognitive neuroscience explains cognition: in terms of neural representations and their neurocomputational processing (Piccinini forthcoming).

RTI conveniently ignores phenomenal consciousness. This is contrary to a long philosophical tradition stretching back at least to Husserlian phenomenology. According to this alternative tradition, a complete account of intentionality must take phenomenal consciousness into account (Horgan and Tienson 2002, Loar 2003, Kriegel 2013, Bourget and Mendelovici 2017). By contrast, RTI theorists hope to explain at least some aspects of intentionality without getting bogged down in the consciousness morass. This essay is an implicit argument that progress can be made by developing RTI further; I leave sorting out the relation between intentionality and consciousness to future work.

In the rest of this essay, I focus solely on *descriptive* mental states and the descriptive (or indicative) representations that correspond to them. I use “descriptive” in a broad sense, according to which an agent may bear a descriptive attitude that P without being committed to P—or, equivalently, without fully believing that P. Examples of such descriptive intentional states include believing, imagining, forming a hypothesis, pretending, etc. A full theory of intentionality must also account for desires, intentions, and other nondescriptive mental states. Nondescriptive attitudes lie beyond the scope of this essay. From now on, when I write “representation” and “intentionality,” I mean *descriptive* representation and *descriptive* intentionality.

**3. Natural and Nonnatural Representation**

As Paul Grice (1957) points out, there are two importantly different notions of meaning: natural and nonnatural. I will use a modified version of Grice’s distinction. According to one usage of “meaning,” a sign means that P just in case it raises the probability that P. This is *natural meaning*. For example, smoke naturally means fire in the sense that the presence of smoke raises the probability that there is a fire; dark clouds naturally mean rain in the sense that the presence of dark clouds raises the probability that it’s raining.

If we are lucky, such signs raise the probability of what they mean to 1. If so, they are factive—they entail that what they mean is true. Grice focuses on this kind of case, so he defines natural meaning as factive. In real life, though, many signs raise the probability of what they mean to less than 1. This is the world we live in, and philosophers have to adapt to it. That’s why I stipulate the weaker condition that the natural meaning that P raises the probability that P, without necessarily raising it to 1.

According to another usage of “meaning,” which Grice dubbed *nonnatural meaning*, a sign means that P whether or not it raises the probability that P. For example, an utterance of “c’è un fiore” may or may not raise the probability that there is a flower, yet it still means that there is a flower. Similarly, “c’è un unicorno” does not raise the probability that there is a unicorn at all, yet it still means that there is a unicorn. Natural languages and other arbitrary sign systems are examples of systems whose well-formed expressions have nonnatural meaning.

The notion of meaning is well suited to characterizing public signs, including natural language utterances. But what I’m investigating here is not, primarily, the meaning of natural language. My main target is the semantic content of the representations that, according to RTI, are *expressed* via natural language. Since my topic is representations and *their* semantic content, I need a notion—analogous to meaning—that applies to representations. The standard thing to say is that representations carry *information*.

Accordingly, Andrea Scarantino and I (2010, 2011) introduced a distinction between natural and nonnatural semantic information that parallels Grice’s distinction between natural and nonnatural meaning. A vehicle carries *natural semantic information* that P just in case it raises the probability that P. By contrast, *nonnatural semantic information* that P need *not* raise the probability that P. Given this notion, a vehicle may carry non-natural information that P even though it does not raise the probability that P. Nonnatural semantic information is a broader notion than nonnatural meaning. For nonnatural semantic information may be carried by any information bearers, including representations internal to a cognitive system, whereas nonnatural meaning is restricted to public signs such as linguistic utterances.

Scarantino and I (2010, 2011) used the notion of representation—understood as a state that can misrepresent or get things wrong—to shed light on the notion of nonnatural semantic information. We implied that to carry nonnatural semantic information that P is the same as to represent that P. That was a mistake: it is entirely possible to represent that P while carrying *natural* semantic information that P. This is not to say that carrying natural semantic information is sufficient for being a representation—later we’ll discuss what else is required. The important point is that carrying *non*natural semantic information is just one way of representing; representing and carrying nonnatural semantic information are not the same thing.

Instead of explicating nonnatural semantic information in terms of representation, I will now use the distinction between natural and nonnatural semantic information to introduce a parallel distinction between two types of representation. This will give us the concepts we need to make progress on intentionality. I will call a representation that carries natural semantic information a *natural representation*. By definition, tokening a natural representation that P—e.g., a perceptual state carrying the natural semantic information that P—raises the probability that P. In addition, I will call a representation that carries nonnatural semantic information a *nonnatural representation*. By definition, tokening a nonnatural representation that P—e.g., a representation corresponding to the intentional act of imagining that P—need not raise the probability that P.[[4]](#footnote-4)

A complete account of (descriptive) intentionality in terms of representation requires both natural and nonnatural representations. Perhaps nonnatural representation can be reduced to natural representation somehow; we’ll discuss this possibility later. For now, we should accept that RTI requires nonnatural representation too. The reason is that there is a large class of broadly descriptive mental states and corresponding linguistic expressions whose tokening need not raise the probability of what they represent. Because they need not raise the probability of what they represent, natural representation—which is just representation that raises the probability of what it represents—cannot fully account for intentionality.

For example, imagining that P need not raise the probability that P. Hypothesizing that P need not raise the probability that P. Considering whether P need not raise the probability that P. Dreaming that P need not raise the probability that P. Pretending that P need not raise the probability that P. Heck, even believing that P may or may not raise the probability that P.

Explaining such mental states in terms of representation requires representations whose tokening need not raise the probability of what they represent—which is also what they carry information about. Accordingly, such representations are nonnatural representations, which carry nonnatural semantic information. An adequate RTI needs nonnatural representations in addition to natural ones.

**4. Functional Role**

As we’ve seen, representations have two aspects: semantic content and functional role. For example, according to RTI, a belief that there is a flower is explained by a representational vehicle with an appropriate functional role carrying the semantic content that there is a flower. Functional role, in turn, is the causing of certain effects under certain conditions. As long as one grants that representations can play a suitable functional role within cognitive systems, some version of the argument in the rest of this essay goes through. Nevertheless, it is worth explicating the relevant functional role in more detail.

In this section, I introduce a multilevel, mechanistic, neurocomputational framework and use it to explicate representational functional role. What follows is a brief summary; a more detailed exposition and defense may be found elsewhere (Piccinini forthcoming).

Minds are complex computational, representational, multilevel mechanistic systems. They are constituted by mechanisms with many levels of organization. A mechanism is a structure that performs activities. It is composed by a set of concrete components (sub-structures) that perform sub-activities and are organized so that the activities of the components constitute the activities of the whole mechanism. The functional role of a well-functioning mechanism is the activity it performs under relevant circumstances. For example, when food enters a well-functioning stomach, the stomach digests the food. Each component of a multilevel mechanism is in turn a mechanism, and each mechanism may be embedded in a larger containing mechanism.

Structures (i.e., potential components of a mechanism) and activities constrain one another: a given structure can only perform a limited range of activities, and a given activity can only be performed by a limited range of structures. Thus, functional roles are not separable—not distinct and autonomous from—the structures that perform them. Nevertheless, often the same activity can be performed by different structures, different arrangements of the same structures, or different arrangements of different structures. Because of this, activities are often multiply realizable—that is, different kinds of mechanism can exhibit the same activity.

Some mechanisms and their components perform *teleological functions*, which are a special kind of functional role. In other words, such mechanisms don’t just perform activities simpliciter—they perform (teleological) functions, which are a special kind of activity. Performing their function at the appropriate rate in appropriate circumstances is what they are supposed to do. If they fail to perform their functions at the appropriate rate in the appropriate circumstances, they malfunction. Performing computations and processing representations are teleological functions of certain specialized mechanisms.

Teleological functions are controversial. Some people find them ontologically suspect, others account for them in terms of the evolutionary or selection history of a system (Garson 2016). I prefer a goal-contribution account: a teleological function is a stable contribution to a goal of organisms.

Organisms have biological and non-biological goals. For present purposes, what matters most are the biological goals, which are survival, development, reproduction, and helping others. Any regular contribution to a biological goal of an organism on the part of a biological trait or artifact is a biological (teleological) function of that trait or artifact.

Some functions may be subject to plasticity and learning, meaning that some traits (e.g., the nervous system) may have the function to acquire more specific functions through developmental and cognitive selection processes. For example, a neuronal population may acquire the function of representing the values of a specific environmental variable through a process of biological development and learning (cf. Millikan 1984 on derived proper functions, Garson 2012, and Neander 2017, 21-2). I will leave a full treatment of the relation between functional plasticity and mental representation to another occasion.

Some multilevel mechanisms are computational and representational, that is, they perform computational and representational functions. Such mechanisms range from primitive computing components such as logic gates and perhaps neurons all the way up to complex networks of processors. Each level is composed of lower level mechanisms and partially constitutes higher-level mechanisms, all the way up to the whole system. Higher-level representations and computations are constituted by lower level ones, all the way down to atomic computations performed by primitive computing components.

Computation does not require representation—there are computations defined over vehicles that are not representations because they have no semantic content. Nevertheless, minds do perform computations over representations. That is how they keep track of what is happening around them and react appropriately, which contributes to their survival and inclusive fitness.

A computation is a kind of mechanistic process defined over certain degrees of freedom of the vehicles being processed, regardless of any more specific physical properties of the vehicles (Piccinini 2015, Chap. 7). Because of this, computations are not only multiply realizable, like many other activities and functions are—they are also medium-independent. In principle, computations may be realized using any medium with the appropriate degrees of freedom. (In practice, there are other constraints to consider, such as temporal constraints on how fast an organism needs to respond to stimuli.)

The kinds of computational systems we are interested in, such as minds, employ *structural* representations. Structural representations are internal states that are systematically related to one another as well as to what they represent. They guide action by modeling, or simulating, what they represent (Craik 1943). For example, a smartphone navigation app helps you find your way, in part, by modeling your environment. Thus, the functional role of a structural representation is guiding action with respect to an environment by simulating the environment or a portion thereof. More precisely, structural representations “stand in” for what they represent by being homomorphic (i.e., partially isomorphic) to their target and by guiding action with respect to it; in short, they are functioning homomorphisms (Gallistel 1990, 2008, Gallistel and King 2009). Contrary to what is sometimes claimed (Ramsey 2007), this notion of representation is the one best suited to make sense of neural representation.[[5]](#footnote-5)

Minds may be natural, and hence biological, or artificial. The same framework I sketch here applies to both natural and artificial minds, even though the mechanisms involved may differ. Since my primary target is human minds and other natural minds, I will set artificial minds aside and focus on natural minds.

The natural minds we are acquainted with are realized in neurocognitive systems, which are systems of multilevel neurocognitive mechanisms. These systems are embodied and embedded in the innocuous sense that they are tightly coupled with a body and environment and their functioning is deeply affected, in real time, by such a coupling. As we shall see, representations acquire semantic content thanks to the coupling between neurocognitive systems and their environment. Perhaps biological minds also include aspects of the body and environment beyond the neurocognitive system as their parts; whether they do makes no difference to the framework I adopt so I will not comment on that any further.[[6]](#footnote-6)

**5. Informational Teleosemantics**

In the previous section, I sketched an account of the functional role of (structural) representation: to be a functioning homomorphism of its target or, equivalently, to be a simulation that guides action. What remains to figure out is how individual representations within a structural representational system acquire semantic content and what content they have.

As a first step, I will briefly sketch a version of Informational Teleosemantics.[[7]](#footnote-7) In later sections, I will extend the theory to cover nonnatural representations. Informational Teleosemantics and related theories have been developed by a number of philosophers over the last few decades (e.g., Stampe 1977; Drestke 1988; Fodor 1987, 1990a, 2008; Millikan 1984, 1993; Ryder 2004, forthcoming; Neander 2017; this literature is surveyed in Adams and Aizawa 2010, Neander 2012, and Neander 2017, Chap. 4). I don’t have space to compare and contrast my version of Informational Teleosemantics with other theories; that’s ok because the differences are mostly irrelevant here. For present purposes, all theories in this family share most of the same strengths and weakness.

Informational teleosemantics has two main ingredients: natural semantic information and (teleological) function. As we have seen, a state (or signal) carries natural semantic information that P just in case it raises the probability that P. The core of Informational Teleosemantics is Natural Representation:

 (NR)

A state (or signal) R occurring within a system S *naturally* *represents* that P =def A function of S is tracking that P by producing R.

A system S *tracks* that P by producing R =def S produces R if and only if P so as to guide the organism in response to the fact that P.

In other words, a state R represents that P just in case the system that produces R has the function of producing R just in case P obtains, in order to guide the organism in response to the fact that P. That is the system’s *representational* *function*. A state R will co-vary with the fact that P to the degree that the system fulfills its representational function.

If the system always fulfills its representational functions, the occurrence of R will raise the probability that P to 1 and thereby be factive. More realistically, systems don’t always fulfill their representational function, in part because the world is full of noise. Therefore, there is often a tradeoff between false positives and false negatives. The more the system works to avoid false negatives, the more it ends up with false positives—and vice versa. Under realistic circumstances, if the system fulfills its representational function enough of the time to generate a reliable enough correlation between R and the fact that P, then R will raise the probability that P to a corresponding degree. To that degree, R carries natural semantic information that P. This is the connection between representation and natural semantic information.

In general, R will be one among a range of similar states (e.g., the firing of a neuronal population at different rates) all of which have the function of tracking specific values of an environmental variable (e.g., the orientation of a line within a specific portion of the visual field). That may include values that the organism has never encountered. That is, a representational system may learn to track, say, oriented lines in the visual environments; when learning is sufficiently advanced, the system may be able to track lines with *any* orientation, including orientations that were never encountered during the learning period. In addition, typical representational states are holistic in the following sense: the parts of a representation mutually affect one another as the system constructs a representation of a whole perceptual field. Taken together, the natural representational states that a system produces track the actual environment. In other words, they constitute a structural representation that simulates the organism’s environment. Thus, this is an account of the semantic content of natural structural representations. For simplicity, however, (NR) abstracts and idealizes these complexities away to focus on individual states in isolation from one another and from their context.

The tracking function mentioned in (NR) amounts to not only producing R—which carries natural semantic information that P—just in case P, but also being appropriately connected with motor control systems so as to guide the agent with respect to the fact that P. One reason to include action guidance is to identify the (distal) fact that P—as opposed to one of its more proximal intermediaries—as the content of a representation.[[8]](#footnote-8)

If a fact P obtains and a representation R is produced and appropriately connected with motor control systems, the system fulfills its representational function by correctly representing that P via R. If P is *not* the case but R is produced anyway, and R is connected with motor control systems so as to guide them, then R *mis*represents that P. If R is produced but fails to be appropriately connected with motor control systems so as to guide the agent with respect to the fact that P, then R fails to represent that P—in fact, it fails to represent anything.[[9]](#footnote-9)

For example: neurons *whose function* is firing in the presence of cows may occasionally fire in the presence of horses, perhaps because the dark night makes it difficult to distinguish cows from horses (Fodor 1990a); in that case, such neuronal firings are still carrying natural semantic information that there is a cow, because in general such firings raise the probability that there are cows. In addition, the organism responds to such firings by making action plans as if there were cows; therefore, the function of such firings is to track cows. If such firings occur in the presence of a horse, they are now misrepresenting the presence of a horse as the presence of a cow. This is how Informational Teleosemantics solves this aspect of the so-called disjunction problem: do those neurons naturally represent the disjunction *cow-or-horse-on-a-dark-night*? No, they represent only cows, because their function is tracking cows.

Many sensory, perceptual, and perceptual-belief states in the nervous system fit this informational teleosemantic account of the semantic content of representation. Representational systems have the function of producing states that co-vary with states in the organism’s body and environments—that is, of producing internal states just in case certain external states occur—so as to guide the organism with respect to those external states. When internal states carry such information and are connected to the organism’s motor control system in the right way, they fulfill their representational function. Otherwise, they misrepresent or fail to represent.

Informational Teleosemantics has many virtues and it explains a very important, basic notion of representation: *natural representation*. It does *not* work for nonnatural representation.

**6. Why Informational Teleosemantics Fails for Nonnatural Representation**

Nonnatural representations are representations that carry nonnatural semantic information, which in turn is information that need not raise the probability of what it’s about. They are the kind of state that, according to RTI, explains intentional states such as imagining, hypothesizing, pretending, entertaining possibilities, and the like. In addition, any instance of believing that P that does not raise the probability that P is a nonnatural representation. Nonnatural representations are also the kind of state that, according to RTI, causes typical natural language utterances and tokens of other arbitrary sign systems.

There is no direct way to reduce nonnatural representation to natural representation. This is because the content of natural representation is natural semantic information, and natural semantic information that P consists in raising the probability that P. Since nonnatural representation that P is defined as a representation that need *not* raise the probability that P, the content of a nonnatural representation that P is, by definition, something other than the natural information that P.

At this point, an optimistic informational teleosemanticist might hope to account for nonnatural representation by adding to (NR) a clause that employs the notion of nonnatural information:

 (NNR)

A state (or signal) R occurring within a system S *nonnaturally* *represents* that P =def

(1) R carries nonnatural semantic information that P and

(2) A function of S is tracking that P by producing R.

Clause (2) still attributes the representational system the function of tracking that P. But nonnatural representations need not have a tracking function. For example, when an agent is imagining that P or entertaining the possibility that P, the function of its representational system is not at all to track that P. Therefore, clause (2) must go.

Can it be substituted with a clause that appeals to the function of carrying nonnatural semantic information? Consider this proposal:

 (NNR\*)

A state (or signal) R occurring within a system S *nonnaturally* *represents* that P =def A function of S is carrying nonnatural semantic information that P so as to guide the organism in response to the nonnatural semantic information that P.

(NNR\*) attempts to mirror (NR) as closely as possible by substituting nonnatural semantic information for natural semantic information. Unfortunately, (NNR\*) has two fatal problems.

First, unlike (NR), (NNR\*) is not a naturalistic account of representational content. Unlike natural semantic information, which raises the probability that something is the case, we have not been given a naturalistic account of nonnatural information. All we know is that nonnatural semantic information need not raise the probability that P. Thus, nonnatural information is not a suitable ingredient for a naturalistic account—at least not until it is naturalized.

Second, (NNR\*) does not lead to a viable account of misrepresentation. (NR) accounts for natural misrepresentation as a kind of malfunction: a natural representation that P misrepresents just in case it is produced when P, which the system has the function of tracking, is not the case. Thus, a natural representation misrepresents just in case the system fails to fulfill its representational function. This does not work for nonnatural representation.

Whether a nonnatural representation misrepresents does not depend on whether it malfunctions. Conversely, whether nonnaturally misrepresenting is a malfunction does not depend on whether the system fulfills its function of carrying nonnatural information. To see this, consider an example.

Suppose that, due to Cotard’s delusion, I believe I’m dead. By RTI, I believe that I’m dead by tokening a mental representation whose content is that I’m dead. If I were really dead, I couldn’t do this. Therefore, my representation that I’m dead cannot track that I’m dead. Therefore, it cannot be a natural representation—it’s a nonnatural representation. I’m not dead but my representational system nonnaturally misrepresents me as dead. By hypothesis, my belief is a delusion due to neurocognitive malfunction. Thus, my belief may well count as a nonnatural representation that misrepresents due to system malfunction.

But now suppose that my neurocognitive system is working just fine. I *know* I’m alive. I just *imagine* being dead, perhaps as part of a meditative practice. As before, by RTI my state of imagining is explained by my tokening a mental representation whose content is that I’m dead. Since I’m not dead, my imaginative state *mis*represents me as dead. Since my representational system cannot track that I’m dead, this is again a case of nonnatural representation. But there is no sense in which my representational system is malfunctioning. If its function is to carry the nonnatural information that I’m dead in the service of guiding action, then it is fulfilling its function. My representational system is functioning perfectly well—there is no malfunction at all.

The point generalizes beyond this somewhat idiosyncratic example. Take any nonnatural representation that P. Under many conditions, the functions of a system that is representing nonnaturally are such that the system ought to represent that P even though P is not the case. It may be entertaining hypotheses, evaluating counterfactuals, or daydreaming. Regardless of the exact condition, a system that is representing nonnaturally may have the function of representing that P whether or not P is the case. If so, then nonnatural misrepresentation is not representational malfunction. Therefore, the standard Informational Teleosemantics account of misrepresentation as representational malfunction does not work for nonnatural representation. And without a viable account of nonnatural misrepresentation, we lack a viable account of nonnatural representation.

In summary, Informational Teleosemantics fails for nonnatural representation for two reasons: nonnatural information has not been naturalized and no viable account of nonnatural misrepresentation has been given.

**7. Offline Simulation**

To construct a viable account of nonnatural representation, I will build on some basic features of structural representation. As we’ve seen, I am assuming that a representational system constructs internal simulations of what it represents—that is, dynamical models that guide action. Let’s consider the case of an organism responding to its environment in real time. The semantic content of its perceptual representational system is its current internal and external environment. For example, the semantic content of the visual system is the visual scene surrounding the organism; the semantic content of the proprioceptive system is the state of the organism’s body.

Environments change constantly, both on their own and in response to organisms’ actions. Therefore, in order to guide an organism successfully, it isn’t enough for a representational system to simulate how the environment is at a given time; such a simulation must be continuously *updated* in real time. In addition, such a simulation cannot be based solely on incoming sensory information about the environment at the time it impinges on the organism, for two reasons.

First, receiving and processing sensory information takes time. Any simulation that is based solely on what sensory information indicates about the present state of the environment will lag behind the state of the environment and be out of date. As the environment evolves, the organism needs a simulation of how the environment is right now, not how it was when it impinged on the organism.

Second, action must be planned in advance. Even real-time responses to the current environment take time to be executed, so they may have to be programmed in response to how the environment will be at the (slightly future) time of action rather than how the environment is right now. For example, walkers must predict where their feet will meet the ground to program appropriate muscle contractions in response to contact with the ground. If they have the wrong expectation, they may lose their balance. The time difference between the present and when action takes place may be small, but under many circumstances successful action requires adjusting for even a small difference.

In short, a representational system must be able to predict how the environment is when a simulation is running, which is slightly later than the state of the environment indicated by incoming information, as well as when action is executed, which is slightly later than now. I will call the difference between the time when action is executed and the time when sensory information impinges on the organism the *sensorimotor time interval*. Because the sensorimotor time interval must be taken into account in order to maintain relatively accurate simulations of the actual environment in real time, internal models must be coupled to their target through a combination of sensory information and extrapolation.[[10]](#footnote-10)

So far, we are still talking about natural representation. Except that we’ve now built into it a fact often neglected in the Informational Teleosemantics literature: in order to function, natural representation must represent more than what sensory experience indicates about the environment at the time the environment impinges on the organism. Natural representation must be based on a combination of sensory information and extrapolation that simulates how the environment is—either right now or in the immediate future.

A representational system may represent more than what current sensory information indicates in many other ways:

1. It may track objects and their location when they are partially or wholly occluded, and hence not perceivable. Obstacles may partially or wholly occlude objects, yet a sufficiently powerful representational system will continue to track the existence and location of some occluded objects. Object completion, object permanence, and cognitive maps are just some classic phenomena involving such tracking.
2. It may keep a trace of past events (episodic memory).
3. It may predict future states of the organism and its environment beyond the sensorimotor time interval, as when a predator intercepts a prey based on where it expects the prey to be at a future time rather than where the prey is now (or within the sensorimotor time interval).
4. It may infer unobservable causes of observable events.

These are still natural representations, or extended forms thereof, as long as their function is to track the actual environment—that is, to co-vary with the environment in order to guide action. Any misrepresentation is a representational malfunction. Nevertheless, these representations depart from standard examples of natural representations by representing more than what is indicated by incoming sensory information.

Paradigmatic examples of natural representations are kept up to date by checking them against incoming sensory information from the environmental variables that are being tracked. By contrast, the kinds of simulations in the above list are maintained in spite of the absence of incoming sensory information from the environmental variables that are being tracked. This shows that natural representation requires the ability to simulate environments in ways that go beyond—are partially decoupled from—incoming sensory information.

I call a simulation that is coupled to its current environment through incoming sensory information an *online* simulation. To the degree that a simulation represents more than what sensory information indicates about the current environment, I call it an *offline* simulation. I have argued that even natural representation involves a degree of offline simulation of the environment.

Offline simulation comes in kinds and degrees. That is, there are different ways that a simulation can be decoupled from incoming sensory information. Each of these ways varies along a continuum:

1. The target of an offline simulation can be more or less extended in space. An offline simulation may involve as little as making the best estimate of what the environment is like where we have our blind spot or as much as forming a cognitive map, complete with landmarks, of an environment we cannot see at the moment.
2. The target of an offline simulation can be more or less extended in time. An offline simulation may involve as little as predicting how the environment will be in a few milliseconds, when our actions will be executed, or as much as anticipating where a prey will be a few seconds from now. When we add the kind of time-tracking and planning abilities that human beings have, it may involve planning actions years in advance as well as remembering events that happened years ago.
3. An offline simulation may be more or less coordinated with online simulation. It may involve as little as filling in a few missing details from an actual visual scene (such as what’s in the blind spot or what lies behind occluding objects) or as much as imagining entire counterfactual scenarios.
4. An offline simulation may require more or less interaction between perceptual systems and motor systems. On the purely perceptual side, it may involve simply filling in visual or auditory details beyond what the incoming sensory information indicates. On the purely motor side, it may involve simply selecting an action that is affordable within the currently perceived environment. Perceptual imagination and motor imagination can also be integrated. A representation of a hypothetical action may require a representation of a currently invisible, hypothetical, or future environment, and it can be fed back to the perceptual imagination system to predict how the environment would change in response, and then perhaps predict how other agents would respond further, and so on, so as to assess the pros and cons of various possible actions before choosing one.
5. An offline simulation may be more or less automatic. It may be as automatic (and unconscious) as filling in the blind spot or as deliberate as planning a vacation.

Some of the above ways and degrees of offline simulation have to do with responding to the current environment in real time. Others do not—they have to do with understanding and communicating past events, planning long-term action, and other cognitive functions that are not directly tied to ongoing (nonlinguistic) action. To understand the ways that offline simulation can augment the representational power of a cognitive system, we must distinguish three importantly different cases.

*7.1 Augmented Online Simulation*

Let’s begin with what I call augmented online simulation: the simulation of the current environment in real time, to guide current action. As we’ve seen, even simulations that are largely online represent more than what is indicated by incoming sensory signals about the state of the environment at the time it impinges on the system. Accordingly, augmented online simulation is built on some combination of incoming sensory information and extrapolation.

Some of the ways that the simulation goes beyond sensory information need not be tracked by the system. For instance, the system need not track that it’s extrapolating the current or immediately future state of the environment or filling in the blind spot. It’s all part of building the most accurate and actionable model of the current environment and using it to guide action.

Other ways in which the simulation goes beyond sensory information must be tracked. An obvious case is object completion: a good visual system represents objects in its environment as partially occluded three-dimensional wholes, not as two-dimensional patches. To do this, the system must track which parts of which objects are visible and which parts are occluded, either by the object itself or by other objects. Such tracking, which is necessary to respond appropriately to the environment, is still a form of natural representation because its function is to track aspects of the current environment.

*7.2 Offline Simulation of the Actual Environment*

Next, let’s consider the offline simulation of portions of the actual environment that cannot be currently perceived. This involves remembering past events, considering what is present in portions of the environment that are not currently perceivable (e.g., non-visible parts of a maze), or predicting future states of the environment for medium- and long-term planning. In addition to tracking the environment itself, such offline simulation requires tracking *that* the simulation is offline. The system must track which aspects of a given simulation are online and which are offline, on pain of getting confused and responding to a non-present, non-actionable portion of the environment.

Again, offline simulation and the tracking of its offline status are a form of natural representation, for their function is to track some aspect of the actual environment. But offline tracking and the tracking of its offline status are not closely coupled with the state of the environment via sensory information in the way that online simulation is. Since the simulation is entirely offline, it cannot be constantly updated and checked for accuracy by using sensory information. Therefore, it constitutes a nontrivial extension of (augmented) online natural representation.

One way to see that this is a nontrivial extension of natural representation is to consider the degree to which an offline misrepresentation of the current environment counts as a malfunction. If an offline representation represents the angle between two branches of a maze as being 90 degrees when in fact it’s 60 degrees, it’s tempting to say that the system is malfunctioning. But it may be more accurate to say that it’s providing the best estimate of the angle based on available natural semantic information. Offline simulation provides the best estimate of an environment that the organism is capable of. As long as the organism is not neglecting more accurate natural semantic information it possesses, the representational system may be functioning correctly even if it doesn’t represent everything completely accurately. [[11]](#footnote-11)

*7.3 Offline Simulation of Nonactual Environments*

Finally, let’s consider the simulation of a nonactual environment. Since the environment is nonactual, it *cannot* be simulated online—any such simulation must be conducted offline. Simulation of a nonactual environment involves forming representations that, according to RTI, explain acts such as imagining counterfactual scenarios, entertaining possibilities and hypotheses, and daydreaming.

A special case is dreaming simpliciter. During a dream, typically the subject experiences nonactual events without tracking that they are nonactual. This does not matter because dreaming occurs while the organism is asleep. While the organism is asleep, normally, the motor system is paralyzed—except for eye movements. While dreaming, motor commands may be issued in response to the experienced stimuli, but such motor commands do not leave the nervous system, so the body does not move—except in pathological cases such as sleepwalking.

In other cases of offline simulation of nonactual environments, which occur while the organism is awake, it is critical that the representational system track that such simulations are offline and the way they depart from the actual environment, on pain of acting in response to the imagined environment rather than the actual one. To illustrate, suppose a hungry organism fantasizes about having a succulent meal in front of it. If it attempts to eat such a meal, it will fail. If it represents the meal as being somewhere in its non-visible environment without representing it as a fantasy, the organism will start searching for such a meal. If it systematically acts in response to such fantasies at the expense of acting on accurate representations of the actual environment, it will eventually starve. To act effectively, the organism’s representational system must keep track that its imaginary meal is not in the actual environment. It must track which of its representations are offline simulations and the way they depart from the actual environment.

To perform such tracking, the representational system must possess some internal state or signal whose function is tracking the degree to which a simulation is offline. Since the function of this state or signal is tracking a current aspect of the (internal) environment, this is still a form of natural representation—it tracks one current aspect of the representational system itself.[[12]](#footnote-12) Yet it provides a crucial ingredient by which the representational system achieves nonnatural representation.

The most important point is that offline simulations of nonactual environments do not fit the standard Informational Teleosemantics story, because they are not coupled with the environment via sensory information. Where do offline simulations of nonactual environments get their semantic content, and what content do they have?

**8. Offline Semantics**

To give an account of the semantic content of offline simulations, we need to say a bit more about what mental simulations consist in—what their parts are and how they are bound together to form representations of objects and their properties. I assume that something like the mainstream view of neural representation is correct: mental simulations consist of the *bound* firing of different neuronal populations at appropriate rates; each portion of the simulation has the function of tracking one aspect of what is represented. For instance, the firing of different neuronal populations at certain rates may represent certain shapes, textures, colors, positions in space, directions of motion, etc. When a combination of different neuronal populations is firing together and their firing is bound via an appropriate mechanism—e.g., their firing is synchronized—the result is a simulation of an object and its properties. This is in line with our understanding of neural representation from mainstream neuroscience (Thomson and Piccinini 2018). Whether the details are correct matters less than that something along these lines is on the right track.

Having said what a mental simulation in general consists in, I assume that *offline* simulations consist in the offline deployment of representational resources—neuronal populations whose firing is bound—that are acquired in the course of online simulations. There is empirical evidence that this assumption is correct, but I will not review such evidence here (for a start, see Barsalou 1999, Kosslyn et al. 2006).

When a representational resource is deployed online, in the service of (augmented) online simulation of the current environment, its accuracy is maintained and updated in light of incoming sensory information. Standard Informational Teleosemantics (NR) applies. That resource represents what it has the function to track in the actual environment.

When a representational resource is deployed offline, in the service of offline simulations of the actual environment, its accuracy can be checked at least in principle by collecting sensory information under appropriate circumstances. An extension of Informational Teleosemantics (NR) still applies. That resource still represents what it has the function to track in the actual environment.

Finally, when a representational resource is deployed offline, in the service of offline simulations of a nonactual environment, everything changes. What is represented is no longer actual, so the simulation is likely to be inaccurate relative to the actual environment. Worse, the actual environment can no longer be used directly to attribute semantic content to the representational resource, so Informational Teleosemantics (NR) does not apply. But the representational resource *could* be used in the service of online simulation. If it were so used, it would have semantic content thanks to standard Informational Teleosemantics.

I propose that when a representational resource is deployed offline, in the service of offline simulations of nonactual environments, it retains the same semantic content it has when it is deployed online. Thus, offline simulations represent what they *would* represent according to Informational Teleosemantics if they were deployed within online simulations. Thus, the content of offline simulations piggybacks on the content of online simulations.

Typical representations consist of many representational resources bound together in the relevant way. Accordingly, a principle of compositionality applies: the content of the whole representation is a function of the content of the representational resources that compose it and the way they are bound together. Here is a simplified example: suppose that a representational system constructs a representation consisting of a HORN representation, a HORSE representation, and the two are bound to one another as well as to a CONTIGUITY representation such that what is represented by the HORN representation is spatially related to what is represented by the HORSE representation in the relevant way. The result is a UNICORN representation, which is a composite representation deriving its content from the content of its constituent representational resources. Representational resources can be deployed together whether or not their *targets* can go together. When targets cannot go together but we represent them together, we represent the impossible.

There is much more to be said about representational resources, how they can be combined, and how they give rise to various aspects of representation. Given the scope of this essay, I have to stop here and return to nonnatural representation.

**9. Nonnatural Representation as Offline Simulation**

We can finally account for nonnatural representation and misrepresentation, and then use nonnatural representation to explain nonnatural meaning.

The first step is to show that offline simulations of nonactual environments represent nonnaturally. By definition, offline simulations of nonactual environments need not have the function of tracking things as they are. If anything, they often have the function representing things as *different* than they are, although their precise function varies from case to case. Now suppose that an offline simulation of a nonactual environment has the function of representing things as different than they are. In order to fulfill such a function, the organism must track that the simulation is offline and the way it departs from the actual environment.

Returning to an earlier example, suppose I intentionally imagine myself as dead, knowing full well that I’m alive. By RTI, I can do this by running a simulation of myself that binds to a simulation of a state of death, while my representational system tracks that representing myself as dead is a departure from reality. My simulation represents myself as dead because it is composed of two bound representational resources that represent both me and the state of death, which they represent because that’s what they represent when they are run online. But this is no malfunction: my simulation is functioning correctly, because it misrepresents myself as dead while the system tracks that this is a departure from reality. The simulation does not raise the probability that I’m dead at all—its existence is actually incompatible with being a correct representation. Therefore, it represents nonnaturally.

Alternatively, suppose I have Cotard’s delusion and imagine myself as dead by running the same sort of simulation as before, while my representational system *fails* to track that representing myself as dead is a departure from reality. I actually believe I’m dead. Now my simulation is both misrepresenting and malfunctioning. In combination with the previous example, this shows that in the case of offline simulations of nonactual environments, misrepresentation and malfunction are logically independent. As I argued in Section 6, this is a characteristic feature of nonnatural representation.

More generally, one of the main functions of offline simulations of nonactual environments is representing things as different than they are in ways that the representational system tracks. When the system successfully tracks the ways in which a simulation departs from (what the system knows about) the actual environment, the representational system functions correctly—it misrepresents because that’s what it’s supposed to do. By contrast, when the system fails to track the ways in which the simulation departs from the actual environment, the representational system malfunctions. Either way, this is a case of nonnatural representation. Its content is determined thus:

(NNR\*\*)

A state (or signal) R occurring within system S *nonnaturally* *represents* that P =def

1. R is deployed within an offline simulation of a nonactual environment.
2. A function of S is tracking the ways in which R is an offline simulation of a nonactual environment.
3. If R were deployed during an online simulation, R would naturally represent that P.

We finally have an account of both the content of nonnatural representation, their function, and the relation between nonnatural representational function and misrepresentation. Clause (1) is entailed by (2) so it’s redundant; I include it for clarity and explicitness. Recall that a state R naturally represents that P just in case it’s produced by a system that has a function of tracking that P by producing R. If a system produces R to fulfill some other function, while having at least the function of tracking the way in which R departs from the actual environment, then R still represents that P, but nonnaturally so. This is the case whether or not the system fulfills its function of tracking the way in which R departs from the actual environment.[[13]](#footnote-13)

If R represents an impossible situation, or a situation that is incompatible with R being produced (such as that I’m dead), (NNR\*\*) applies to the representational resources that constitute R. That is, each representational resource that constitutes R represents what it would represent if it were deployed during an online simulation, and R represents what the bound representational resources that constitutes it would represent if they could be deployed during an online simulation.

(NNR\*\*) is an account of the semantic content of nonnatural mental representations, which is a step towards a naturalistic account of intentionality. The next challenge would be to cash out the mentalist theory of meaning by explaining the nonnatural meaning of linguistic utterances in terms of NNR\*\*. The best I can do here is sketch how such an explanation would go.

Before we get to the kind of nonnatural representation that explains linguistic intentionality and nonnatural meaning, we need one more ingredient. We need an arbitrary signaling system—a system of signals that are arbitrarily assigned to their targets. Notice that the kind of representational resources we’ve been working with—the ones that compose into mental simulations—are not arbitrary. Natural representations are acquired and deployed in response to specific stimuli in order to guide action with respect to those stimuli. Their non-arbitrariness is also the reason they have the semantic content they have. Their format is not separable from their content.[[14]](#footnote-14)

But organisms need to communicate, and to communicate effectively they need public signals. Public signals can be associated with referents relatively arbitrarily and can become an effective means of communication so long as organisms have a way to coordinate their use of the signals (Barrett and Skyrms 2017, Skyrms 2010).

Arbitrary signaling systems, by themselves, are not enough to give rise to nonnatural meaning. For an arbitrary signaling system may be deployed solely to express natural representations. Consider vervet monkeys. They have a set of alarm calls for indicating to other monkeys when predators are around. They use different signals for different types of predator. As far as I know, the function of alarm calls is to track and communicate the presence of predators. If this is their only function, then vervet monkey alarm calls have only natural meaning, even though they are an arbitrary signaling system.

When arbitrary signaling systems are combined with offline simulations of nonactual environments, however, they acquire the kind of nonnatural meaning that explains linguistic intentionality. If a linguistic utterance U expresses the force and semantic content of mental state R, and the representational system has the function of tracking the actual environment by producing R, then the function of U may still be to track the actual environment. If a linguistic utterance U expresses the force and semantic content of mental state R, and the representational system produces R while fulfilling a non-tracking descriptive function—e.g., the function of misrepresenting how things are—then the function of U is to fulfill the relevant non-tracking descriptive function. Since utterances may or may not have the function of representing correctly, they may or may not raise the probability of what they describe. Therefore, their semantic content is nonnatural (rather than natural) meaning.

Needless to say, this is just a bare sketch of an account of linguistic intentionality. Many complexities are sidestepped. At a minimum, a more adequate account would describe the ways in which simulations with different functional roles and contents give rise to utterances with different forces and meanings.

**10. Conclusion**

I have argued that an adequate explanation of intentionality needs to account for nonnatural meaning and nonnatural representation, namely, meaning and representation that need not raise the probability of what they represent. I have also argued that mainstream accounts of the semantic content of mental representations—Informational Teleosemantics and related theories—fail for nonnatural representation and a fortiori cannot account for nonnatural meaning.

One reason is that nonnatural misrepresentation—as opposed to natural misrepresentation—is not representational malfunction. Unlike the natural representations that traditional representational theories of intentionality tend to posit and account for, the function of a nonnatural representation is not necessarily to accurately represent the actual state of an environmental variable. Rather, it may be to represent the state of an environmental variable inaccurately or even to represent something that is not in the environment at all.

For example, if I want to imagine that I am dead, the function of my representation is to represent that I am dead, and my representational system would malfunction if it failed to represent me as dead. Here is another example: if I want to deceive someone by lying, what would actually be a malfunction of my cognitive-linguistic system is telling the truth. These are examples of misrepresentations whose function is precisely to misrepresent![[15]](#footnote-15)

Thus, malfunction (simpliciter) cannot explain the distinction between a nonnatural representation that represents correctly and a nonnatural representation that represents incorrectly. That is, malfunction (simpliciter) cannot explain misrepresentation by nonnatural representations. A fortiori, malfunction (simpliciter) cannot explain nonnatural representation.

I have sketched an account of nonnatural representation (and misrepresentation) in terms of natural representation plus offline simulation of nonactual environments plus tracking the ways in which a simulation departs from the actual environment. Nonnatural representation plus an arbitrary sign system gives rise to nonnatural meaning.

To represent nonnaturally, the system must be able to decouple internal simulations from sensory information by activating representational resources offline. The system must be able to represent things that are not in the actual environment and to track that it’s doing so, i.e., there must be an internal signal or state that can indicate whether what is represented departs from the actual environment. In addition, the system must be able to manipulate a representation independently of what happens in the actual environment and keep track that it’s doing so. The simplest case might be a system that manipulates an internal model based on potential actions without carrying out the action. This does not require an imaginative system separate from the motor system; it only requires the ability to manipulate parts of the motor system offline and redirect an efferent copy of its output to the perceptual system, while keeping track that the system is offline. To summarize: nonnatural representations are offline simulations whose departure from the actual environment the system has the function to keep track of.

The follow up question is, what makes an offline simulation, which does not correlate with anything in the perceivable environment, represent something? I proposed that these states represent based on a compositional semantics that begins with representational resources that can be deployed online and represent what they would represent if they were deployed online. In other words, nonnatural representational content piggybacks on natural representational content. As in mainstream Informational Teleosemantics, natural representations get their content from *performing the function* of tracking a specific environmental variable. Nonnatural representations get their content by deploying natural representational resources to produce representations that do not track, and often lack the function of tracking, the actual environment. So nonnatural representations always represent what their natural counterparts have the function of representing.

With this framework in place, we are in a position to explain the difference between a state that misrepresents by mistake (malfunctioning natural representation) and a state that, at least in some cases, misrepresents on purpose (correctly functioning nonnatural representation). Notice that nonnatural representations may also misrepresent by mistake, for instance in cases of delusion. The organism is mistakenly nonnaturally representing that P and believing that P, when in fact P is not the case. When an organism is wrong in this way, she is failing to track the decoupling between her internal models and the actual environment.

The organism naturally represents both its environment and its actions. Representation of the environment can be decoupled from immediate sensory stimuli or even from the actual environment. Such representations of the environment can be manipulated. As long as the organism correctly tracks that she is manipulating her own mental simulations decoupled from the actual environment (for planning purposes, for entertainment purposes, or to deceive someone), she can represent the environment incorrectly while fulfilling the system’s representational function. When she fails to keep track of her own interventions on her own representations, she makes a mistake.

So nonnatural representation is the manipulation of natural representational resources decoupled from the actual environment. Functional nonnatural representation is nonnatural representation in which the organism keeps track of her representational manipulations. Dysfunctional nonnatural representation is nonnatural representation in which the organism fails to keep track of her sensory decoupling and representational manipulations.

This proposal breaks new ground and makes substantive progress towards an adequate naturalistic account of intentionality in terms of mental simulations implemented by neural representations. A lot more must be worked out before a full account of intentionality is at hand. For starters, the above account of nonnatural representation must be combined with an arbitrary signaling system to account for the nonnatural meaning of natural language utterances.

**References**

Adams, F. and Aizawa, K., "Causal Theories of Mental Content", *The Stanford Encyclopedia of Philosophy*(Spring 2010 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2010/entries/content-causal/>.

Barrett, J. A., and Skyrms, B. (2017). Self-assembling Games, *The British Journal for the Philosophy of Science*, 68 (2): 329–353.

Bartels, A. (2006). Defending the structural concept of representation, Theoria, 21(55): 7-19.

Barsalou, L.W. (1999). Perceptual symbol systems. Behavioral and Brain Sciences, 22, 577–660.

Bechtel, W. (2008). Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience. London: Routledge.

Bourget, David and Mendelovici, Angela, "Phenomenal Intentionality", *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.), forthcoming URL = <https://plato.stanford.edu/archives/spr2017/entries/phenomenal-intentionality/>.

Clark, A. (2013). “Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science.” Behavioral and Brain Sciences, 36(3): 181-253.

Churchland, P.M 2012. Plato's Camera: How the Physical Brain Captures a Landscape of Abstract Universals. Cambridge, MA: MIT Press.

Craik, K. (1943). The Nature of Explanation. Cambridge: Cambridge University Press.

Cummins, R. (1996). Representations, Targets, and Attitudes. Cambridge, MA: MIT Press.

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

Dretske, F. (1986). Misrepresentation. In 'Belief: Form, Content, and Function', edited by R. Bogdan. Oxford: Clarendon (pp. 157-173).

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

Fodor, J.A. (1987). Psychosemantics: The Problem of Meaning in the Philosophy of Mind. Cambridge, MA: MIT Press.

Fodor, J.A. (1990a). A Theory of Content and Other Essays. Cambridge, MA: MIT Press.

Fodor, J.A. (1990b). “Information and Representation.” In Philip P. Hanson (ed.), *Information, Language and Cognition*. University of British Columbia Press, pp. 513-524.

Fodor, J.A. (2008). LOT 2: The Language of Thought Revisited. Oxford: Oxford University Press.

Gallistel, C. (1990). Representations in Animal Cognition: An Introduction, Cognition, 37(1-2): 1-22.

Gallistel, C. (2008). Learning and Representation. In 'Learning and Memory: A Comprehensive Reference', edited by J. Byrne. Amsterdam: Elsevier (pp. 227-242).

Gallistel, R. G., & King, A. P. (2009). Memory and the computational brain: Why cognitive science will transform neuroscience. New York: Wiley/Blackwell.

Garson, J. (2012). Function, selection, and construction in the brain. Synthese

189: 451–481.

Garson, J. (2016). A Critical Overview of Biological Functions. Dordrecht: Springer.

Grice, P. (1957). Meaning, Philosophical Review, 66: 377-388.

Grush, R. (2003). In Defense of Some 'Cartesian' Assumptions Concerning the Brain and Its Operation, Biology and Philosophy, 18(1): 53-93.

Grush, R. (2004). The Emulation Theory of Representation: Motor Control, Imagery, and Perception, Behavioral and Brain Sciences, 27(3): 377-396.

Horgan, T., and Tienson, J., 2002, “The Intentionality of Phenomenology and the Phenomenology of Intentionality,” in D. Chalmers, *Philosophy of Mind: Classical and Contemporary Readings*, Oxford: Oxford University Press, pp. 520–933.

Hohwy, J. (2013). The Predictive Mind. Oxford: Oxford University Press.

Isaac, A. (2013). Objective Similarity and Mental Representation, Australasian Journal of Philosophy, 91(4): 683-704.

Jacob, Pierre, "Intentionality", *The Stanford Encyclopedia of Philosophy*(Winter 2014 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2014/entries/intentionality/>.

Kriegel, U., ed. (2013). *Phenomenal Intentionality*, Oxford: Oxford University Press.

Kosslyn, S., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. New York: Oxford University Press.

Loar, B., 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*, Cambridge, MA: MIT Press, 229–258.

Millikan, R. (1984). Language, Thought, and Other Biological Categories. Cambridge, MA: MIT Press.

Millikan, R. (1989). Biosemantics, The Journal of Philosophy, 86(6): 281-297.

Millikan, R. (1993). White Queen Psychology and Other Essays for Alice. Cambridge, MA: MIT Press.

Millikan, R. (2000). Naturalizing intentionality, The Proceedings of the Twentieth World Congress of Philosophy, 9: 83-90.

Morgan, A. (2014). “Representations Gone Mental.” Synthese, 191(2): 213-244.

Morgan, A. and Piccinini, G. (2018). “Towards a Cognitive Neuroscience of Intentionality.” Minds and Machines, 28 (1): 119-139.

Neander, Karen, "Teleological Theories of Mental Content", *The Stanford Encyclopedia of Philosophy*(Spring 2012 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2012/entries/content-teleological/>.

Neander, K. (2017). *A Mark of the Mental: In Defense of Informational Teleosemantics*. Cambridge, MA: MIT Press.

O'Brien, G. and Opie, J. (2004). Notes toward a Structuralist Theory of Mental Representation. In 'Representation in Mind', edited by H. Clapin, P. Staines, and P. Slezac. Amsterdam: Elsevier (pp. 1-20).

Piccinini, G. (2015). *Physical Computation: A Mechanistic Account*, Oxford: Oxford University Press.

Piccinini, G. (forthcoming). *Neurocognitive Mechanisms: Explaining Biological Cognition*. Oxford: Oxford University Press.

Piccinini, G. and Scarantino, A. (2011). “Information Processing, Computation, and Cognition. *Journal of Biological Physics*, 37.1, pp. 1-38.

Pitt, D. (2017). "Mental Representation", *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.), forthcoming URL = <https://plato.stanford.edu/archives/spr2017/entries/mental-representation/>.

Ramsey, W. (2007). Representation Reconsidered. Cambridge: Cambridge University Press.

Robbins, P, and Aydede, M., 2009, *The Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press.

Ryder, D. (2004). SINBAD Neurosemantics: a theory of mental representation, Mind & Language 19(2): 211-240.

Ryder, D. (forthcoming). *Models in the Brain*. Oxford University Press.

Scarantino, A. and Piccinini, G. (2010). “Information without Truth”. *Metaphilosophy*, 43.3, pp. 313-330.

Shagrir, O. (2012). Structural Representations and the Brain, The British Journal for the Philosophy of Science, 63(3): 519-545.

Shepard, R. and Chipman, S. (1970). Second-Order Isomorphism of Internal Representations: Shapes of States, Cognitive Psychology, 1(1): 1-17.

Skyrms, B. (2010). Signals: Evolution, Learning and Information. Oxford: Oxford University Press.

Speaks, Jeff, "Theories of Meaning", *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.), forthcoming URL = <https://plato.stanford.edu/archives/spr2017/entries/meaning/>.

Stampe, Dennis. 1977. Toward a causal theory of linguistic representation. In Midwest Studies in Philosophy: Studies in the Philosophy of Language. vol. 2., ed. Peter A. French, Theodore E. Uehling Jr., and Howard K. Wettstein, 81–102. Minneapolis: University of Minnesota Press.

Swoyer, C. (1991). Structural Representation and Surrogative Reasoning, *Synthese*, 87(3): 449-508.

Thomson, E., and Piccinini, G. (2018). Neural Representations Observed, *Minds and Machines* 28 (1): 191-235.

Waskan, J. (2006). *Models and Cognition: Prediction and Explanation in Everyday Life and in Science*. Cambridge, MA: The MIT Press.

1. Thanks to Michael Barkasi, Alex Morgan, Andrea Scarantino, Joulia Smortchkova, Alberto Voltolini, and an anonymous referee for helpful comments on previous versions. For helpful feedback on presentations that led to this essay, thanks to audiences at the University of Turin, University of Missouri – St. Louis, “Mental Representation: The foundation of cognitive science?” Bochum, Germany, September 2015 and “Information and its Role in Science: Physics, Biology, Cognitive and Brain Sciences,” Jerusalem and Tel Aviv, Israel, June 2016. Thanks to Alex Schumm for editorial assistance. This material is based upon work supported by the National Science Foundation under grant no. SES-1654982. [↑](#footnote-ref-1)
2. See Morgan and Piccinini 2018 for an opinionated literature review, which paves the way for the present project. [↑](#footnote-ref-2)
3. RTI is my label for that portion of the Representational Theory of Mind that addresses intentionality. [↑](#footnote-ref-3)
4. I call it “nonnatural” only because that is the term Grice used. I am not implying that nonnatural representation cannot be naturalized. On the contrary, by the end of this paper I will sketch a naturalistic account of nonnatural representation. [↑](#footnote-ref-4)
5. For other expositions of the notion of structural representation and, in some cases, for arguments that it’s the notion best suited for neuroscience, with some variation in the details, see, among others, Shepard and Chipman 1970; Swoyer 1991; Cummins 1996; Grush 2003, 2004; O'Brien and Opie 2004; Ryder 2004, forthcoming; Bartels 2006; Waskan 2006; Bechtel 2008; Churchland 2012; Shagrir 2012; Isaac 2013; Hohwy 2013; Clark 2013; Morgan 2014; Neander 2017, Chap. 8. Some of these authors also point out that structural representations simulate what they represent, that simulations may be run offline, and that running simulations offline increases the representational power of a system. As far as I can tell, none of them develops offline simulation into an account of nonnatural representation, as I do below. [↑](#footnote-ref-5)
6. For more on whether and how the mind is embodied and embedded, see Robbins and Aydede 2009. [↑](#footnote-ref-6)
7. Some proponents of structural representation take it to provide an *alternative* to Informational Teleosemantics by accounting for semantic content in terms of homomorphism (or structural similarity) instead of information. But homomorphism, in order to amount to representational content, must be established and sustained through a causal or informational connection with a referent. That’s where Informational Teleosemantics comes in. Conversely, when a system of internal states systematically carries information about a system of external states, it is homomorphic to it. Thus, homomorphism and information are two sides of the same semantic coin (cf. Morgan 2014). [↑](#footnote-ref-7)
8. For a more detailed discussion of the problem of distal content, including a solution along similar lines, see Neander 2017, Chap. 9. [↑](#footnote-ref-8)
9. Motor control systems may be dysfunctional or disconnected from the locomotive system, as when the organism is paralyzed. Whether that is the case is a separate question from whether the system represents things. Paralysis does not prevent representation. [↑](#footnote-ref-9)
10. This point is reminiscent of some Bayesian approaches to neuroscience and psychology (e.g., Hohwy 2013, Clark 2013), except that I remain neutral on the exact balance between sensory information and prediction that is at play. As far as I am concerned, the balance between sensory information and prediction may vary depending on circumstances. [↑](#footnote-ref-10)
11. Another way to see that this is a nontrivial extension of natural representation is that this extension is not even possible under causal theories of representational content, which posit that the content of a representation is its present cause (e.g., Fodor 1990a, Neander 2017). In general, the content of an offline simulation of the actual environment is not its present cause, and certainly it need not be its present cause (cf. Fodor 1990b for an argument to this effect as well as precursors to some of the present themes). By contrast, (NR) does not require that the semantic content of a representation be its present cause, so it supports the extension from augmented online simulations to offline simulations of the actual environment. [↑](#footnote-ref-11)
12. That’s not to say that such internal tracking is metarepresentational. It may simply amount to constructing simulations that play an appropriately modified role in guiding action whenever representational resources are deployed in a way that departs from what the system represents to be the actual state of the world. [↑](#footnote-ref-12)
13. Morgan (2014) argues that being a structural representational system, and even being a structural representational system that conducts offline simulations, is insufficient for being a mental representation. His reason is that some plants may possess structural representational systems that conduct offline simulations (e.g., of the day-night cycle). If plants don’t have a mind, then, there must be more to mental representation than just offline simulation. One feature of mental representations that plants lack is the offline simulations *of nonactual environments*. There are probably others, such as medium independence (Piccinini forthcoming). [↑](#footnote-ref-13)
14. I suspect that this is why mental states, which are at least in part constituted by mental representations, have original rather than derived intentionality. Original intentionality is intentionality that does not derive from any other intentionality. It is widely believed that natural language utterances and other tokens of public sign systems derive their intentionality from the intentionality of mental states. Mental states, in turn, are believed to have original intentionality. [↑](#footnote-ref-14)
15. Batesian mimicry is an example of signals that may be seen as carrying nonnatural information without being produced by a representational system in the present sense. Species that are harmful to predators often send warning signals—such as specific sounds or coloration patterns—to discourage predators from attacking them. Such warning signals carry the natural information that the would-be prey is harmful. In Batesian mimicry, a harmless species imitates the warning signals that a harmful species produces. From the perspective of the predator, the mimic’s warning signal might still carry natural information in the sense that a mimic’s warning signal raises the probability that a potential prey is harmful (Skyrms 2010, 75-7). From the perspective of the mimic, however, the function of its warning signals is to *mislead* predators—that is, to cause predators to *mis*represent the mimic as harmful. In this sense, a mimic’s warning signals carry nonnatural information. [↑](#footnote-ref-15)