ISSN 2299-0356

*Filozoficzne Aspekty Genezy* — **2022, t. 19, nr 1**

*Philosophical Aspects of Origin* s. 133–163

https://doi.org/10.53763/fag.2022.19.1.198

ARTYKUŁ ORYGINALNY / ORIGINAL ARTICLE

Alexander Rosenberg

Duke University

# How to be an Eliminativist

Received: March 9, 2022. Accepted: June 21, 2022. Published online: July 21, 2022.

Abstract: In the 40 years since its first promulgation, contemporary eliminativism about intentional content has secured considerable additional support in the form of both neuroscientific findings and an absence of significant counter-evidence within the now greatly expanded study of the brain and its components. This paper reports some of the most telling of these results. Three serious issues remain to be dealt with by philosophical proponents of eliminativism: claims that neuroscience's frequent use of the word "representation" requires or presupposes that neural circuitry actually carries such content, claims that the phenomenology of first-person introspection reveals the undeniable existence of intentional content, and arguments to the effect that eliminativism is self-refuting, contradictory or pragmatically paradoxical, owing to its claim that there are no true assertions. This paper addresses these three arguments against eliminativism.

Keywords:

eliminativism;
intentionality;
neuroscience;
representation;
consciousness;
self-refutation argument;
theories of truth

## 1. Introduction

Eliminative materialists deny that there are beliefs and desires (and other propositional attitudes) in the brain (or anywhere else, for that matter). This thesis was originally based on arguments about the explanatory weakness of theories that attribute intentional content to the brain or its components. [1] Since then, ad-

[1] See Paul M. Churchland, "Eliminative Materialism and the Propositional Attitudes", *Journal of Philosophy* 1981, Vol. 78, No. 2, pp. 67–90; Patricia S. Churchland, **Neurophilosophy: Toward a Unified Science of the Mind/Brain**, MIT Press, Cambridge 1986.

vances in neuroscience have considerably strengthened eliminativism by furnishing detailed evidence of how the brain and its components actually do work to deliver behavior. Section 1 reports some of these findings. However, philosophers and others continue to resist eliminativism, mainly for three unrelated reasons: some conjecture that models and theories in neuroscience report how brain states *represent*, and that representation is intentional; other philosophers also argue that first-person introspection makes it undisputable that consciousness has intentional content, so that the existence of intentionality cannot be denied without rejecting the thesis that cognitive agents are (sometimes) conscious; finally, many philosophers accept the view that eliminativism is incoherent or plagued by a pragmatic paradox, since it defends a self-referential and self-refuting thesis: one whose intentional content is that there is no intentional content. Sections 2, 3 and 4 of the present paper address these three challenges to eliminativism. The last section treats the real challenges facing eliminative materialism as having the form of a philosophical thesis.

## 2. Considerations in Support of Eliminativism

Eliminativists reject what used to be called "folk psychology", together with its more recent development, the theory of mind. As it figures in social psychology, the latter is something of a more explicit version or formulation of the explanatory theory all normal *Homo sapiens* employ to explain and predict their own behavior and that of other humans, along with many other vertebrates that engage in environmentally appropriate behavior. One way to articulate the theory of mind as it is employed in social psychology is given by the "boxology" in Figure 1 below.

The boxology can't express one crucial feature of the theory of mind's causal claims: the way beliefs and desires pair up in the mind to bring about choices, decisions and actions is via the match-up, the relevance, of their *propositional contents* to one another. Of the indefinitely many beliefs and desires in a subject's head, the ones that find their way into the belief box and the desire box in the figure below do so owing to the relevance of their contents to one another. In the simplest case, the contents of the belief box express the means whereby the content of the desire box can be attained. Thus, the contents that determine the character of the subject's behavior have a semantics, a meaning, that is at least sometimes accessible to the subject. The causal role of the belief that Paris is the capital

of France differs from the causal role of the belief that 2 is the only even prime owing to the differences in their propositional contents.

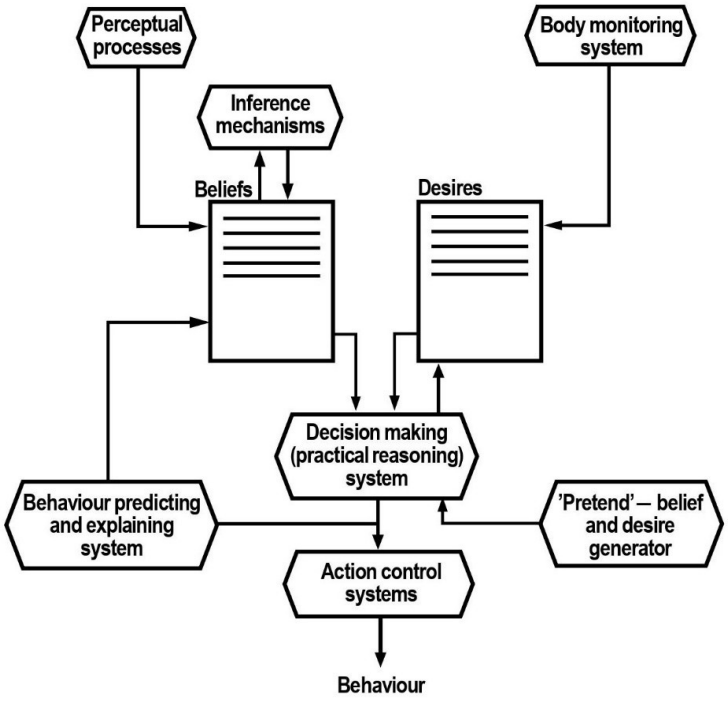



Figure 1. [2]

Initial reasons to reject the theory of mind included predictive weakness in its intended domain of application — normal human decision and choice, and the limitations on its ability to explain abnormal human behavior. [3]

Eliminativists recognize that there was a "cup-half-full/cup-half-empty" type of disagreement in play here.

[2] Reprinted, with permission, from: Shaun Nichols, Stephen Stich, Alan Leslie, and David Klein, "Varieties of Off-Line Stimulation", in: Peter Carruthers and Peter K. Smith (eds.), **Theories of Theories of Mind**, Cambridge University Press, Cambridge 1996, p. 40 [39–74].

[3] See Churchland, "Eliminative Materialism and the Propositional Attitudes…"; Churchland, **Neurophilosophy…**; Stephen Stich, **From Folk Psychology to Cognitive Science**, MIT Press, Cambridge 1983; Stephen Stich, "Do True Believers Exist? A Reply to Andy Clark", *Aristotelian Society Supplement* 1991, Vol. 65, pp. 229–244.

The cup-half-full: It's obvious that human affairs have been arranged in accordance with the theory of mind since time immemorial. Human cultural, social, political and legal institutions have been built on the assumption that humans are responsible for their behavior and that this responsibility is the result of the normal operation of packages of beliefs and desires that drive the behavior. The earliest literary works known to us employ this theory to give meaning to their narratives. They reflect the likelihood that the theory of mind has been applied in largely the same form probably since humans acquired language. Its local success in predicting the behavior of *small* numbers of collaborators and competitors in our *immediate vicinity* over *short* periods of time is literally unrivaled. For it has no rival. Its local predictive success in the Pleistocene and the absence of rivals in all subsequent human history underwrote the ever-increasing explanatory employment of the theory of mind, outward from its origin to explain and predict along three distinct dimensions: increases in the number of agents, increases in their spatial distance from the user of the theory, and increases in their temporal distance in terms of both earlier and later. The theory of mind's predictive power gets weaker and weaker as the numbers of people increase (too many to watch), as their distances from the user of the theory of mind increases (they are out of sight), and as time periods lengthen away from the instant of the theory's employment (in more distant pasts and farther futures). But since it had no rival, the theory of mind's predictive failures did not undermine its explanatory use.

The cup-half-empty: One tipoff to eliminativists that the theory of mind's explanatory/predictive cup is half empty is the fact that the theory's domain of predictive success has remained unchanged both in precision and in range over the millennia since it began to be employed. Another is that the theory has not been improved, either by increasing the precision of its explanatory variables, or by the identification of systematic interfering (*ceteris paribus*) factors, or by the discovery of operational measures for its causal variables, over the same period, of millennia, during which it has been employed. [4]

[4] A glance at the literature of behavioral economics is enough to show that even self-consciously scientific, laboratory-driven, systematic approaches aimed at improving rational choice theory (the theory of mind formalized) have failed to either enhance quantitative prediction or increase explanatory precision of the theory. See Michael Joffe, "Mechanism in Behavioral Economics", *Journal of Economic Methodology* 2019, Vol. 26, No. 3, pp. 228–242; Nathan Berg and Gerd Gigerenzer, "As-If Behavioral Economics: Neoclassical Economics in Disguise?", *History of Economic Ideas* 2010, Vol. 18, No. 1, pp. 133–165.

Failure to improve in respect of predictive range, and failure to improve as regards predictive precision, over the longest time period available, are signal marks to eliminativists of explanatory impoverishment. An explanatory theory that is on the right track should at least show some predictive improvement in range and precision over the several (hundreds of) thousands of years it has been in use. Accordingly, eliminativists hold, there is something seriously wrong with the theory of mind.

Even so, in the absence of a rival theory with at least some hope of improving on the precision and the range of the theory of mind, there is little incentive to surrender it. Indeed, when we add in the apparent obviousness to introspection of the truth of the theory of mind, the idea of surrendering it begins to seem laughable.

Eliminativism about belief/desire psychology becomes much more attractive when a positive rival theory of behavior becomes available. This is just what has happened: recent developments in neuroscience have shifted the balance of the arguments in favor of eliminativism away from the largely philosophical and methodological to the factual, experimental and empirical.

Cognitive neuroscience is beginning to explain in detail how human (and other mammalian) brains deliver behavior. Most striking have been the advances in understanding how the brain delivers behavior that the theory of mind purports to explain. The mechanism that has so far been uncovered is nothing like what the theory of mind tells us it should be. Not only is there nothing in recent discoveries by neuroscience that would vindicate the boxology of the theory of mind, but also the actual mechanism of how information [5] is acquired, stored and deployed in the brain to direct behavior reveals that there is no scope even for the kinds of causal variables that the theory of mind posits — let alone causes that pair up in virtue of semantic content.

The research program that unraveled the theory of mind can be traced from the first experiments in the 1950s on HM, the patient famous for being unable to form a wide variety explicit and declarative beliefs owing to destruction of his

[5] Of course, "information" must be interpreted here as a notion free from intentionality: for example, as Shannon and Weaver employ it to indicate probability reduction. See Section 5 below, where I discuss Brian Skyrms, **Signals: Evolution, Learning and Information**, Oxford University Press, New York 2010.

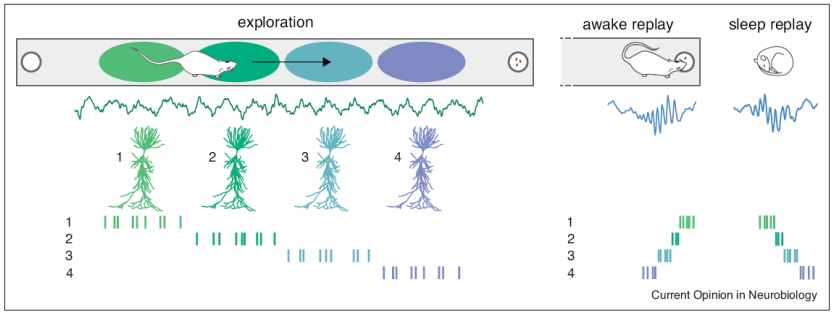
hippocampus. Inspired by these findings, Kandel undertook to identify the macromolecular construction of implicit, and then explicit, declarative beliefs in the brain, and in particular in the hippocampus. He found it, and the work was rewarded with the Nobel Prize in 2000. This inspired the further discoveries of O'Keefe and the Mosers, who shared the 2014 Nobel Prize for their work. Neuroscience had finally zeroed in on how the brain acquires, stores and deploys the information that the theory of mind mistakenly describes as beliefs, and that the theory of mind mistakenly describes as desires. Here we will focus on beliefs: in particular, explicit beliefs about the local environment and one's place in it.

HM's inability to form and retain many kinds of new explicit, propositional beliefs was traced to the destruction of his hippocampus in a medial temporal lobe ablation. This brain-structure and its immediately surrounding tissue — the entorhinal cortex, became the locus for studies that undermine the theory of mind as an explanation of behavior (including behaviors that in humans would be described as "action"). Work on rats, primates and humans eventually identified the neural circuits (in the entorhinal cortex — the grid cells and other specialized neurons) that carry environmental information as regards local geography, the subject's location and direction, and the presence of threats and rewards. Neuroscientists can locate different clusters of cells that fire depending on the shape and size of a lab animal's space. The neurons do not map the space in any sense. There is no physical isomorphism between them and the space they map. It is just different packages of neurons firing exactly the same pulses depending on the size of the space the animal finds itself in. Neuroscientists can "read" the space's dimensions and topology off of which neurons fire. Similarly, they can detect the animal's direction, speed, and other explicit information it "has" about the environment from firing in other entorhinal neurons.

Neuroscientific research has located the cells in the hippocampus (of rats, primates and humans) — the place cells — where this information is combined. They have identified the simple algorithm that combines the grid cells' oscillating action potentials into dampening and strengthening superpositions in the place cells. By identifying just the order and the strength of individual neuronal cell firings, neuroscientists can identify the animal's location, direction and future path. Differences in information about environment and behavior are all just matters of firing sequence and strength of action-potentials in neural circuitry. Even the long-term storage of specific information about the environment in the prefrontal

cortex consists in the hundred-fold temporal compression of the same sequences with the same strength, sometimes in the forward direction, sometimes reversed. Returning to the hippocampus from the prefrontal cortex, these compressed firing sequences are decompressed to combine in superpositions with neural sequences in the nucleus accumbens. The latter cells store the result of reinforcement and punishment neural conditioning, to determine actual behavior. By reading off the firing of the neural circuits in the brain, and without knowing anything about the subject's previous experience, the neuroscientist can accurately predict the subject's behavior, including what common sense would describe as choice. (No intentional stance is required.) It is worth emphasizing that the information the neural circuitry stores is not encoded in some "morse code" that the neuroscientist needs to decrypt. Positional information in the brains of experimental subjects is only a matter of the temporal order and strength in which neurons discharge their electrochemical potentials. Oscillations from neurons in the entorhinal cortex that record local geography, the subject's direction and speed, are combined at "place cells" in the hippocampus to produce oscillations that locate the subject. All the neuroscientist must do in order to know the subject's location and direction of travel is read off which place cells are firing. All the neuroscientist has to do to predict how the subject will choose between alternative paths is identify oscillation patterns coming back to the hippocampus from the prefrontal cortex. [6]

[6] The illustration below compactly illustrates several of the discoveries that reveal the non-intentional character of information storage that the theory of mind characterizes as explicit geographical beliefs. All the work is done by sequence and strength of neural firing. Source (with permission): https://tiny.pl/93n6z [02.03.2022].




Sharp wave-ripple (SPWR)-associated neuronal replay. When a rat is running on a linear track, the hippocampus oscillates at theta frequency (green trace) and place cells are successively activated as the rat enters their respective place fields (colored ellipses), yielding a neuronal sequence (vertical ticks). Upon arrival at the food well, the rat stops to consume a reward and the place cell sequence reactivates in reverse order (reverse replay) during SPWRs (blue trace). In subsequent slow wave sleep and quiet rest periods, place cells reactivate in the same order as during exploration (forward replay).

In respect of the sort of information that can be attributed with precision to laboratory animals such as rats, the actual events, states and processes in the brain show nothing remotely like beliefs (or desires for that matter). In particular, they have none of the intentionality, the content, the aboutness, the representational character that is of the essence of propositional attitudes. The actual neural processes in the brain do exactly what beliefs do for the subject, without being anything like beliefs.

If nature makes no jumps, if human brains operate the way rats' brains do in respect of what the theory of mind identifies as their explicit beliefs about their local environments, then there are no propositional attitudes in the human brain either. The homologies in mammalian brain structure, right down to the neural circuitry and the individual neurons, are enough to give neuroscientists confidence that humans acquire, store and deploy information about their environments in the same way rats do. It is just that we have 86 billion neurons to the rat's 21 million: all of the same types as rat-neurons, and arranged in a topography quite similar to neurons in the rat's brain.

---

For details, see Edvard I. Moser, Yasser Roudi, Menno P. Witter, Clifford Kentros, Tobias Bonhoeffer, and May-Britt Moser, "Grid Cells and Cortical Representation", *Nature Reviews Neuroscience* 2014, Vol. 15, pp. 466–481; Edvard I. Moser, "Grid Cells and the Entorhinal Map of Space", Nobel Lecture 2014, December 7, https://tiny.pl/93n6b [02.03.2022]; May-Britt Moser, "Grid Cells, Place Cells, and Memory", Nobel Lecture 2014, December 7, https://tiny.pl/93nvq [02.03.2022]; John O'Keefe, "Spatial Cells in the Hippocampal Formation", Nobel Lecture 2014, December 7, https://tiny.pl/93nvm [02.03.2022]; John O'Keefe and Jonathan Dostrovsky, "The Hippocampus as a Spatial Map: Preliminary Evidence from Unit Activity in the Freely-Moving Rat", *Brain Research* 1971, Vol. 34, No. 1, pp. 171–175; Amir H. Azizi, Laurenz Wiskott, and Sen Cheng, "A Computational Model for Preplay in the Hippocampus", *Frontiers of Computational Neuroscience* 2013, Vol. 7, article number: 161, https://doi.org/10.3389/fncom.2013.00161; George Dragoi, "Internal Operations in the Hippocampus: Single Cell and Ensemble Temporal Coding", *Frontiers in Systems Neuroscience* 2013, Vol. 7, article number: 46, https://doi.org/10.3389/fnsys.2013.00046; Eric R. Kandel, "The Molecular Biology of Memory Storage: A Dialog between Genes and Synapses", Nobel Lecture 2000, December 8, https://tiny.pl/93nvv [02.03.2022]; John L. Kubie and Steven E. Fox, "Do the Spatial Frequencies of Grid Cells Mold the Firing Fields of Place Cells?", *Proceedings of the National Academy of Sciences USA* 2015, Vol. 112, No. 13, pp. 3860–3861, https://doi.org/10.1073/pnas.1503155112; Jai Y. Yu and Loren M. Frank, "Hippocampal-Cortical Interaction in Decision Making", *Neurobiology of Learning and Memory* 2015, Vol. 117, pp. 34–41, https://tiny.pl/93nbq [02.03.2022]; Jai Y. Yu, Kenneth Kay, Daniel F. Liu, Irene Grossrubatscher, Adrianna Loback, Marielena Sosa, Jason E. Chung, Mattias P. Karlsson, Margaret C. Larkin, and Loren M. Frank, "Distinct Hippocampal-Cortical Memory Representations for Experiences Associated with Movement *versus* Immobility", *eLife* 2017, Vol. 6, e27621, https://doi.org/10.7554/eLife.27621.

## 3. "Representation" in Neuroscience

The eliminativist implications of contemporary neuroscience are obscured most thoroughly, even among neuroscientists themselves, by their use of a single word: "representation". Neuroscience employs the word "representation" in many of the models formulated and employed to explain behavior. Elsewhere, and especially among philosophers, this word is employed to identify propositional attitudes identified by the facts, states of affairs, and propositions, that the sentences they contain, or other tokens, are "about". But, as we'll see, neuroscience employs this word via a radical redefinition that deprives "representation" of the meaning that it carries in the theory of mind.

It's worth noting that, in the context of the theory of mind, the word "representation" emphasizes the content-bearing character of beliefs and desires. They *re*-present things and states of affairs beyond, outside, independent of the subject, that are present and presented to it by the environment (or seem to be so presented). The beliefs and desires in which the theory of mind trades are individuated, distinguished, identified by, and given their causal powers by these re-presentations that they "contain". Recall how the theory of mind tells us which beliefs and desires pair up in the mind to bring about choices, decisions and actions: via the match-up, the relevance of their contents to one another. These contents that determine the character of the subject's behavior have a semantics, a meaning accessible to the subject that represents objects of desire and means relevant to their attainment.

Nothing like this happens in the brain: there are no representations with these semantic features in the brain, at any level of organization. When neuroscientists employ the word "representation", they mean something quite different from what that word connotes in the theory of mind. The use of the word "representation" in neuroscience obscures the character of its models and so confers the illusion that intentionality obtains in the brain.

The intentionality-free work that "representation" actually does is the subject of an important book, Nicholas Shea's (Lakatos-award winning) **Representation in Cognitive Science**.[7] Shea notes that "we now have a wealth of empirical data

[7] See Nicholas Shea, **Representation in Cognitive Science**, Oxford University Press, Oxford 2018.

against which to formulate and test theories of neural representation". [8] He draws some of his most detailed examples from the Nobel Prize winning work described above, on the way in which the hippocampus and entorhinal cortex record, store and deploy information about spatial location and other features of local environments. [9] However, Shea works with several other examples of how representation figures in neuroscientists' models. [10]

Examining the research literature of cognitive neuroscience, Shea extracts what he describes as two sufficient conditions for "representation" as employed in cognitive neuroscience: one invokes correlation and the other structural isomorphism. The analysis begins with a characterization of task-function familiar from etiological or selected effects accounts of "function". A task-function is a process with a stabilized outcome. Stabilized outcomes come in three disjunctive kinds: first, the familiar hard-wired Darwinian adaptative outcomes; second, adaptive outcomes that result from learning; and third, ones that contribute to the organism's persistence. When such processes are robust (usually as a result of fine-tuning by feedback), they constitute task-functions. Some task-functions proceed by operating algorithmically, [11] syntactically, [12] over internal physical components [13] of the organism that bear "exploitable relations" to the organism's external environment. [14] Exploitable relations are ones that are actually employed to discharge the task function; they are ones that causally explain the stability and

[8] Shea, **Representation in Cognitive Science...**, p. 27.

[9] See Shea, **Representation in Cognitive Science...**, p. 113–116.

[10] See Matthew F.S. Rushworth, Maryann P. Noonan, Erie D. Boorman, Mark E. Walton, and Timothy E. Behrens, "Frontal Cortex and Reward Guided Learning and Decision Making", *Neuron* 2009, Vol. 70, No. 6, pp. 1054–1069; John K. Kruschke, "ALCOVE: An Exemplar Based Connectionist Model of Category Learning", *Psychological Review* 1992, Vol. 99, No. 1, pp. 22–44; Valerio Mante, David Sussillo, Krishna V. Shenoy, and William T. Newsome, "Context-Dependent Computation by Recurrent Dynamics in Prefrontal Cortex", *Nature* 2013, Vol. 503, pp. 78–84; David C. Van Essen and Jack L. Gallant, "Neural Mechanisms of Form and Motion Processing in the Primate Visual System", *Neuron* 1994, Vol. 13, No. 1, pp. 1–10; Quentin J.M. Huys, Neir Eshel, Elizabeth O'Nions, Luke Sheridan, Peter Dayan, and Jonathan P. Roiser, "Bonsai Trees in Your Head: How the Pavlovian System Sculpts Goal-Directed Choices by Pruning Decision Trees", *PLOS Computational Biology* 2015, Vol. 8, No. 3, e1002410.

[11] See Shea, **Representation in Cognitive Science...**, p. 36.

[12] See Shea, **Representation in Cognitive Science...**, p. 39.

[13] See Shea, **Representation in Cognitive Science...**, p. 32.

[14] See Shea, **Representation in Cognitive Science...**, p. 35.

robustness of the process. The internal components constitute representations when they bear exploitable relations to the environment in one of two different ways.

The first of these two exploitable relations is *correlation* between properties of the internal physical components and properties of external environmental items. To be a *representation* the correlation must be algorithmically treated in ways that causally explain the *stability and robustness* of the task-function. The internal physical components that satisfy these requirements are one of the two kinds of contentful *representations* — as the expression is employed in neuroscience.

The second and distinct sufficient condition for representation is the existence of a structure among the internal physical components that is *physically isomorphic* with items in the organism's external environment, so that the isomorphism, the correspondence, is actually [15] employed as an exploitable relation. [16] The centrality for neural representations of algorithmic, purely structural manipulation over physically defined tokens is something Shea emphasizes persistently:

> [I]nternal processing over components standing in exploitable relations to features of the environment can amount to the implementation of an algorithm, an algorithm by which the system performs various input-output mappings. […] [I]f we take a relevant input-output mapping, content is fixed by the exploitable relations carried by components which make the internal processing an implementation of an algorithm by which the system instantiates that mapping […] task functions give the input-output mappings that are relevant to content determination. That was because a cluster in which the outcomes stabilized by natural selection, learning or contribution to persistence are also produced robustly and are generated by an algorithm that makes use of exploitable relations. [17]

In what follows, I shall assume that Shea's analysis of the way in which cognitive neuroscience employs the concept of representation is accurate. It should be apparent that neither the components of Shea's analysis, nor the ways in which he puts them together to systematize and underwrite the employment of the term

[15] See Shea, **Representation in Cognitive Science…**, p. 119.

[16] As neuroscientists employ the word, "representations" in the brain may often simultaneously satisfy both of these independent sufficient conditions. Shea explains and illustrates circumstances in which they do so, but shows why the structural correspondence does not reduce to correlation.

[17] Shea, **Representation in Cognitive Science…**, p. 110.

"representation" in cognitive neuroscience, imply that neural states, processes or events have intentionality, aboutness or propositional content.

As Shea himself recognizes, the notion of representation that he has analyzed as being at work in cognitive neuroscience is not the kind invoked in the theory of mind, folk psychology or what is widely labeled "the personal" as opposed to the "sub-personal" level of explanation and description. As several centuries of argument in the philosophy of mind has shown (all the way back to Leibniz' mill), and as Shea himself realizes, the theory of mind — a.k.a. folk psychology — explains and predicts behavior at what is often called "the personal level" (by contrast with the sub-personal one). The theory of mind requires a quite different notion of representation from the one at work in cognitive neuroscience that he has systematized. The widely accepted irreducibility of intentionality to purely physical transactions in the brain reflects the difference between the personal and sub-personal levels. Shea identifies four features of the personal-level notion of intentionality that he sets aside as not relevant to the representations neuroscientists locate in the brain: "I will use the term »sub-personal« to cover representations for which content-determination does not depend on [...] [four] complicating features: consciousness, justification for the person, a role in reason-giving interactions with other people, or being structured like natural-language sentences". [18] Shea's account "disclaim[s] these four complicating factors". His project is not one of upgrading the physical into the mental.

The eliminativist about intentionality will treat Shea's impressive account as tantamount to a tacit admission by neuroscientists that their theories of representation provide no reduction, naturalization, or other explanatory foundation for the theory of mind. Their models of representation certainly do not vindicate the existence of states with those four features: i.e. states of belief and desire. What would?

## 4. Against the Argument from Phenomenal Intentionality

It is obvious and unarguable that everyone *reports* that their conscious thoughts, and especially their occurrent beliefs and desires, have representational content. This conclusion seems to be vouchsafed by the introspective phe-

[18] Shea, **Representation in Cognitive Science...**, p. 26.

nomenology of consciousness: that at least sometimes, in fact almost always, when you have a thought, "it feels like" it has occurrent representational content: while you experience the thought, you can "feel" its representational character — that it has its content is evident in your consciousness of the thought.

Eliminativism need not deny that everyone (including we eliminativists) have this feeling. We deny that introspectively accessible feelings of "what it is like" for our thoughts to represent are grounds to conclude that they actually do so. We go further: there are experiments one can perform on one's own phenomenology that reveal that thoughts by themselves do not have representative content.

There has long been an active research program of philosophers seeking to link intentionality and consciousness. Indeed, intentionality has been called upon to explain the nature of consciousness, especially by exponents of the representational theory of conscious experience. Some philosophers have, additionally, argued that all intentional states are conscious states — phenomenally present states of awareness. [19] Even among those who recognize that much cognition is nonconscious, there are philosophers who assert that such thoughts have some sort of intentionality insofar as they have a disposition to be accessed in consciousness. [20]

Philosophers employing the notion of intentional experience in the brain to elucidate consciousness are, of course, helping themselves to a notion that eliminativists reject. But those who equate intentionality with consciousness advance a thesis that eliminativists must confront, since we do not deny the existence of conscious experience. These philosophers do not merely argue that consciousness provides evidence that thoughts have intentional content: they argue that intentionality consists in consciousness or *vice versa*. On this view, the only way to be an eliminativist about intentionality is to hold that we are all zombies.

The weaker claim that introspection is enough to warrant our knowledge that thought has intentional content requires us to accept as probative phenomenology that cannot be subject to intersubjective observation. Most people, and many

[19] See e.g. Terence E. Horgan and John L. Tienson, "The Intentionality of Phenomenology and the Phenomenology of Intentionality", in: David J. Chalmers (ed.), **Philosophy of Mind: Classical and Contemporary Readings**, Oxford University Press, Oxford 2002, pp. 520–533.

[20] See John R. Searle, **Intentionality: An Essay in the Philosophy of Mind**, Cambridge University Press, Cambridge 1983.

philosophers, are prepared to do so. Some philosophers report some introspectively available feeling or other that they claim is a distinctive immediate non-inferential proprietary mark of content. The label for such a feeling is "cognitive phenomenology". [21] Eliminativists will immediately challenge the notion that there is a qualitative "what it's like" feeling that reliably marks one's experience as a thought directed at an object. After all, they will ask, what is it about a feeling that makes it about something beyond or "outside of" itself? The eliminativist will argue that even if there is such a phenomenally available distinct feeling, the conclusion that it signposts intentional content is a learned "interpretation" of the feeling, on the model of the bodily "aboutness" of pains, whose directedness to the site of injury is learned from other experiences (and behavior), and not directly and immediately given in conscious experience. [22]

Moreover, eliminativists shouldn't even grant that there is such a proprietary feeling component accompanying the alleged content in thought. Perhaps the most psychologically powerful considerations an eliminativist can offer to counter these claims about the phenomenology of thought consist in offering step-by-step instruction in how to undertake phenomenological experiments of the sort that will undermine confidence in the self-evident existence of intentional content in thought. We herewith present such an experiment, which readers can run themselves. It begins with data reported by proponents of the existence of phenomenal intentionality. Horgan and Tienson [23] invite their readers to experience the difference between two different conscious states. They claim that the difference between the states immediately reveals the intentionality of conscious thought. Here is one of their examples. It works only once, the first time you hear the noises produced by an out-loud reading of the following inscription:

dogs dogs dog dog dogs

If you have never been exposed to this noise before you will almost certainly attach no meaning, no content, to the conscious state of auditory stimulation the

[21] See Galen Strawson, "Cognitive Phenomenology: Real Life", in: Tim Bayne and Michelle Montague (eds.), **Cognitive Phenomenology**, Oxford University Press, Oxford 2011, pp. 285–325.

[22] See Dale Jacquette, "Sensation and Intentionality", *Philosophical Studies* 1985, Vol. 47, No. 3, pp. 429–440.

[23] See Horgan and Tienson, "The Intentionality of Phenomenology...".

sounds produce. When it is drawn to your attention that "dog" can serve as a noun and a verb in English (meaning to hunt, track or follow), the next time you hear the same noises, your conscious experience will have a content: roughly, that canines tracked by other canines also track canines. The advocate of phenomenal intentionality invites you to accept that the difference in the two conscious experiences consists in the presence of intentional content in the latter. [24]

Now, the eliminativist invites you to re-analyze the phenomenology experienced and to endorse a different conclusion. When you first hear the noises you have one sequence of mental images or other tokens (this may vary from person to person). On the second occasion the only difference is a quite different sequence of mental images or other tokens that runs through conscious experience. The difference between the two conscious experiences is not intentionality, but just more phenomenology — the new mental imagery.

The reader is invited to try the same strategy with similar inscriptions that have been advanced or may be advanced to isolate the intentional experience of propositional attitudes from their content.

Cows cows cow cow cows.

is another example, as is

Visiting relatives can be boring.

The ideation of the suite of noises produced by reading this sentence aloud will be quite different depending on what mental image or other token is provoked by the noise "visiting". Ask yourself, how does thinking of "visiting" as a noun differ from thinking about that sound as a verb? Is it a distinctive intentionality or aboutness or simply a difference in mental imagery?

It is open to advocates of the existence of phenomenal intentionality to argue that conscious sensory states are intentional, and that therefore the intentionality of propositional attitudes reduces to the intentionality of conscious states of sensory awareness. The trouble with such an argument is that the intentionality of

[24] Two distinct instances, tokens, of the same sequence of (silent) noises in consciousness may be followed by different behaviors, thus leading us or others to attribute different intentional content to each token. But note that the role of behavior in this attribution reveals that content is not intrinsic to the sensory experience, but also requires behavioral sequelae.

purely sensory or perceptual states is arguably "non-conceptual" — free from description in words (even in a language of thought), whereas propositional attitudes are clearly ones that employ mental word-tokens. It might be held that sensory awareness of any kind has *nonconceptual content*. However, the complexities and controversies surrounding the nonconceptual content of sensory awareness hardly lend themselves to any claim that such states provide strong, let alone indispensable, phenomenological evidence for the existence of intentionality, aboutness, or representational content. [25]

Many philosophers who believe that conscious sensory experiences are intentional hold that consciousness consists in, reduces to, or is explainable in terms of intentionality. It is ironic, though, that philosophers who endorse this view cash consciousness in for intentionality and then offer a functional theory of how sensory states have their aboutness in terms of the teleosemantic predecessors of Shea's theory. [26] They altogether share the eliminativists' denial of any explanatory role to a "what it's like" phenomenology of consciousness.

Eliminativists can admit that we are all subject to the phenomenological illusion [27] that thought has intentional content, that the illusion is powerful, and that it can at best only be temporarily counteracted or suspended. In this respect it is quite like other human illusions: for example, that physical objects are colored, impenetrable, smooth or rough, hot or cold, etc. Eliminativists can accept that the illusion that thoughts have content explains a great deal about human life, human artifacts and human institutions, and that it has had considerable adaptive value in human evolution. The theory of mind that embodies the illusion that thought has content was an indispensable solution to a design problem (of coordination and cooperation) that *Homo erectus* faced when it found itself at the bottom of the food chain on the African savanna a million or more years ago. Eliminativists deny none of these things. But they accept that just as natural science eventually revealed the actual nature of physical objects to be utterly different from what our

[25] See Tamar Szabó Gendler and John Hawthorne (eds.), **Perceptual Experience**, Oxford University Press, Oxford 2006.

[26] See e.g. Fred Dretske, **Naturalizing the Mind**, *Bradford Books*, MIT Press, Cambridge 1995.

[27] "Illusion" is another term, like "information" and "representation", that needs to be treated as free from the presupposition of intentional content. It should be treated as describing behavior: in particular, behavior that, in Shea's sense of "representation", is not isomorphic with certain real properties of local environments.

conscious sensory experience of them led us to suppose, it also reveals that thought is completely different from what conscious experience led us to suppose.

## 5. Getting Beyond the Charge of Self-Refutation

The problem of pragmatic contradiction that eliminativism faces is the tip of an iceberg — or, perhaps, the canary in a coal mine: one that signposts a set of fundamental issues in philosophy that eliminativists must take seriously. The eliminativists' research agenda includes problems in the philosophy of language, the philosophy of logic, parts of evolutionary game theory, along with the domains of psycholinguistics and linguistic anthropology that intersect with philosophy. A concise and effective response to the objection from pragmatic contradiction would be desirable, but looking for it threatens to distract us from the serious problems that the nonexistence of propositional attitudes reveals. The program of conceptual revision required by eliminativism is so demanding that only our confidence in what the science has revealed and will reveal about the brain makes it worth undertaking.

To see the dimensions of the program, let us consider two widely accepted theses about meaning. The first is Searle's claim that the intentionality of public language is *derived*: that the intentional content of inscriptions and noises is conferred on them by *original* intentionality — the content of mental acts of symbolic interpretation. [28] The derived/original intentionality distinction grounds public sentences, inscription, noises and other *symbols*' meanings in the intentionality within speakers' or inscribers' heads that gives noises or marks the status of *symbols*.

The second thesis motivating the unintelligibility of eliminativism is due to Grice's insight that a speaker's meaning is a matter of the speaker's having a nested set of desires and beliefs about how his auditors will respond to his spoken noises. [29] If I say "I beg your pardon", my speaker-meaning might be "How

[28] See John R. Searle, "Minds, Brains and Programs", *The Behavioral and Brain Sciences* 1980, Vol. 3, No. 3, p. 424 fn. 2 [417–424]; Searle, **Intentionality...**.

[29] On Grice's formula, a means p by uttering x ≡ a intends in uttering x that (1) his audience come to believe p, (2) that his audience recognize this intention, and (3) that (1) occurs on the basis of (2).

dare you?", whereas the sentence-meaning is "I regret I have offended you". [30] Thus, Grice's theory elucidates a kind of meaning distinct from Searle's: speaker-meaning. Notice that detecting Gricean speaker-meaning is just the application of the theory of mind to verbal/inscriptional behavior.

Arguably, between them these two claims seem to exhaust the kinds of meanings recognized by the philosophy of language. Together they imply not just that when eliminativists utter their claims their speaker-meanings bely their eliminativism: they also imply that eliminativism lacks even the resources to allow for meaningful discourse altogether.

Boghossian saw this consequence clearly enough, and traced the problem eliminativists face to an even more consequential matter:

> [T]he *best* arguments for the claim that nothing mental possesses content would count as *equally* good arguments for the claim that nothing linguistic does. For these arguments have nothing much to do with the items being mental and everything to do with their being *contentful*: they are considerations, of a wholly general character, against the existence of items individuated by content. If successful, then, they should tend to undermine the idea of linguistic content just as much as they threaten its mental counterpart. [31]

For, as Boghossian notes, "the relevant notion of content may be assumed to consist simply in the idea of a truth condition". [32] The eliminativist's argument against beliefs and desires is based on the fact that they are supposed to play their causal role in virtue of their contents, the sentences they contain, and that the very numerical identity of each proposition attitude is constituted by the sentence or statement it contains. These contained statements are true or false. The statements' truth is what makes beliefs true, their falsity is what makes them false. Thus beliefs having content consists in their having truth conditions, being true or false. But vocalizations and inscriptions have truth conditions as well, and so have content, the same sort of content propositional attitudes have — ones given by their truth conditions. Eliminativists must therefore also deny that spoken and

[30] See Paul Grice, "Meaning", in: Paul Grice, **Studies in the Way of Words**, Harvard University Press, Cambridge 1989, pp. 213–223.

[31] Paul A. Boghossian, "The Status of Content", *Philosophical Review* 1990, Vol. 99, No. 2, p. 171 [157–184] [emphases in the original].

[32] Boghossian, "The Status of Content…", p. 174.

written tokens (including their own tokens) have content, i.e. truth conditions.

Add Boghossian to Searle and Grice, and it becomes clear that eliminativists must give us a whole new approach to the nature of language — and, for that matter, to truth and falsity. This obligation has, of course, been recognized by eliminativists as far back as Churchland. [33] The task looked daunting enough four decades ago. But giving an account of the way language works that doesn't rely on attributing a truth-conditional semantics to it no longer appears as intimidating as it did in the 1980s. Moreover, providing an account of how it works by appealing to truth-conditionality no longer seems so straightforward.

The significant achievement of teleosemantics was to recognize the ways in which Darwinian processes, both genetic and ontogenetic, shape the fine-grained functions of cognitive states in controlling behavior. Its program of naturalizing intentional content did not succeed — but, eliminativists concede, it did come close. Eliminativists tendentiously hold that from its beginnings with Dretske, Neander, Papineau and Millikan, all the way through to Shea's analysis of representation in cognitive neuroscience reported above, the teleosemantic program has shed a great deal of light on the purely non-intentional causal processes that produce the *illusion* of content in the brain. *Mutatis mutandis*, a similarly Darwinian approach explains the *illusion* of content in public acts of speech and writing — their illusory appearance of conveying speaker-meaning. It does so by approaching the biologically identifiable functions that stand behind the illusory appearance of sentence-meaning.

As Shea's analysis of representation in the brain emphasizes, task-functions emerge in three different ways: the path trodden by the selection of phenotypic traits, identified by Dawkins as extended phenotypes, [34] meaning behaviors that manifest themselves at the level of the organism and its environment; the shaping of linguistic behavior over development and experience that stabilizes outcomes by learning; and, finally, those that stabilize outcomes that enhance the organism's persistence. The Darwinian processes, operating genetically and culturally over generations, give the features to inscription-types and vocalization types that confer on them and their tokens the illusion of sentence-meaning. These tokens,

[33] See Churchland, "Eliminative Materialism and the Propositional Attitudes…", p. 89.

[34] See Richard Dawkins, **The Extended Phenotype: The Long Reach of the Gene**, Oxford University Press, Oxford 1982.

additionally, convey the illusion of speaker-meaning (in the form of the phenomenological illusion illustrated above), as they stabilize particular outcomes for individuals. Exactly how they do so is roughly the same as the way in which, according to what teleosemantic theories tell us, the environmentally appropriate firing of neural circuity conveys the illusion (to observers like us) of intentionality.

Public human language starts to evolve well after the emergence of the neural circuitry on which teleosemantics focuses. But the foundations from which language evolves may long predate the emergence of anything with the neural complexity that might encourage the attribution of intentional content. We can be confident about the possible evolution of signaling systems by Darwinian processes independent of any neural complexity thanks to an insight of David Lewis [35] (1969) and its elaboration by Brian Skyrms. [36] Lewis' aim was to show, contra Quine, that *convention*s could emerge from non-conventional behavior. Skyrms' work shows how, once conventions are in place, language can evolve from them. He did so by adding Darwinian selection to Lewis' game-theoretical models while subtracting from them conscious human thought, in order to develop a mathematical theory of the evolution of language from non-linguistic behavior. Over the course of several decades, employing both mathematical proof and computer simulations, Skyrms has shown how Lewis' signaling model can emerge by Darwinian selection of random interactions among populations as simple as microorganisms. Darwinian processes will in fact favor the persistence of such signaling, which Skyrms calls *proto-language*. Given repeated interaction across well-defined spatial structures (for example matrices or rings of organisms), randomly generated effects on other organisms can be shaped into signaling systems that will spread within populations by operant learning (Darwinian cultural selection), or across generations by Darwinian genetic selection. Skyrms summarizes it thus:

> We have investigated the evolution of signalling in some modest extensions of Lewis signalling games with multiple senders and receivers. [...] Simple models such as those discussed here can be assembled into more complex and biologically interesting systems. The network topologies themselves may evolve. [...] There are all sorts of interesting variations. [...] But the main business of signalling networks is to facilitate successful collective action. The simple models studied here focus on the crucial as-

[35] See David Lewis, **Convention**, Harvard University Press, Cambridge 1969.

[36] See Skyrms, **Signals...**.

> pects of coordinated action. Information is acquired by the units of the group. It is transmitted to other units and processed in various ways. Extraneous information is discarded. Various kinds of computation and inference are performed. The resulting information is used to guide group decisions that lead to coordinated action. All this can happen either with or without conscious thought. [37]

Modeling in evolutionary game theory thus gives us some confidence that to fulfil its coordination functions, linguistic behavior doesn't need original intentionality, aboutness, propositional content or representation. Eliminativists will continue to exploit a Darwinian approach to the evolution of language underwritten by these results. In particular they will hope to show that what truth-conditional semantics identifies as the referential and predicational components of communication can be understood in terms of their functional role — their selected effects.

Spoken language presumably emerged long before inscriptions, and did so not as symbols but as signs: grunts and gestures employed with the function of coordination and cooperation in shared behavior. There will have been strong Darwinian cultural selection pressure for the emergence of increasingly complex vocables, together with their syntactic arrangements, as the behavior they could coordinate itself became more fitness-enhancing. Full blown linguistic competence of the sort that moved human beings from the bottom of the African savannah food-chain to the top requires a large stock of vocables, and syntactic arrangement. It will also have many spandrels — spin-offs from its fitness-enhancing function that are also subject to cultural selection.

In building the vocabulary of public languages, Darwinian processes "search" for noises that work adaptively in the shaping of behavior. As metaphysicians like Paul have recognized, [38] there is nothing logically inevitable even about so basic a set of linguistic devices as objects and predicates. If these devices are ubiquitous in public languages, it is presumably because they had a Darwinian pedigree. Consider the words that emerge from selection for predicates that label salient

[37] Skyrms, **Signals...**, p. 279. Skyrm's use of information must be non-intentional, since his project is to upgrade non-intentional Shannon and Weaver behavior probability-changing: for instance, the production of sounds and marks — something approaching human language. See Peter Godfrey-Smith, "Review of Brian Skyrms' **Signals**", *Mind* 2012, Vol. 120, No. 480, pp. 1288–1297, for an incisive exposition of Skyrm's project.

[38] See L.A. Paul, "Categorical Priority and Categorical Collapse", *Proceedings of the Aristotelian Society* 2013, Vol. 87, Supplementary Volumes, pp. 89–113.

threats and opportunities. Many of them, especially words that identify features of normal sensory experience, will be at the same time highly adaptive and wildly misleading about their user's ambient environments. Color terms are the obvious examples. It is safe to say that the "manifest image" shared across almost all cultures and civilizations surviving through the Holocene is riddled with these words that are adaptive while seriously defective as descriptions of reality. Getting a handle on their systematic functional role in vocalization and inscription provides a reliable way of identifying their contribution to various types of linguistic expressions. This will also often be enough to grasp the functional contribution which compound tokens of these sounds and inscriptions make to individual verbalizations of the kind that Gricean speaker-meaning is intended to elucidate. Eliminativists should treat speaker-meaning and sentence-meaning as flawed but useful human instruments, whose strengths and weaknesses are explained by a theory that is fully eliminativist in its rejection of intentionality.

It is probably only once predictively accurate scientific theories began to emerge, a million or so years after human communication got started, that we could be confident that any of our predicates picked out real properties of things, instead of illusory adaptations. John Locke called these "primary qualities", and in his "New work for a theory of universals" [39] Lewis called them "universals" and "natural properties". Many of the predicates of ordinary language don't name natural properties, ones with systematic scientific explanatory power. They have all the inductive warrant of Goodman's "Grue" and "Bleen". Some are in even worse shape. Almost all of the predicates employed to describe ordinary material objects suffer from logical infirmities that have motivated the most influential of philosophers to deny that they identify any distinct objects at all. [40] Eliminativists insist that the same fate befalls the propositional attitude predicates. As with the predicates of folk physics and folk biology, those of folk psychology — including, of course, its theory of mind — are destined to be replaced, at least for serious science, by a quite different set of predicates. These will be ones whose predictive and explanatory success more strongly encourages the conclusion that there are real properties "behind them" that carve nature at the joints. When it comes to ro-

[39] See David Lewis, "New Work for a Theory of Universals", *Australasian Journal of Philosophy* 1983, Vol. 61, No. 4, pp. 343–377.

[40] See e.g. Peter Van Inwagen, **Material Beings**, Cornell University Press, Ithaca 1990; Peter Unger, "There Are No Ordinary Things", *Synthese* 1979, Vol. 41, No. 2, pp. 117–154.

dent behavior in the lab this has already happened.

It will take centuries to uncover exactly why and how most of our familiar predicates fail to pick out real properties of objects, states, processes and events, even as they perform their duty of helping humans survive and thrive in our adaptive environments. Meanwhile, nothing shows more powerfully how unreliable even the most confident description based on phenomenology can be than the cognitive neuroscientist's designing of optical illusions. These have been employed by neuroscientists to isolate, separate and study experiences that natural selection and human experience have packaged together into adaptive but provably erroneous packages. The best optical illusions deprive all who experience them of confidence in the reliability of their own phenomenal experiences. [41]

At some point or other in the development of language the vocables that function the way the English predicates "__ is true" and "__ is false" do emerged. At roughly the same time, or more probably later, the referential and predicative apparatus of natural languages also emerged. From a Darwinian point of view, it's pretty easy to see why they were selected for. These vocables were extremely useful in fine-tuning adaptive behavior, especially among speakers with different informational resources coordinating and collaborating in the context of projects where survival was crucial. In the environment of early evolutionary adaptation, at the bottom of the food chain on the African savannah, cooperation, and in particular coordination among individuals scavenging and hunting, would have demanded agreement on tactics and on the environmental factors that shape optimally effective strategies.

Millenia after these vocables acquired their task-functions and became ubiquitous, Plato and Aristotle, among others, presumed to provide explicit accounts of the properties the truth- and falsity-predicates named. It is widely held that it is to them that we owe the correspondence and other realist theories of truth (**Metaphysics**, 1011 b25; **Cratylus**, 385b; **Sophist**, 2636). Correspondence theories of truth, despite their appeal over the next two millennia, have never been free from controversy among philosophers. These controversies have not spilled over from academic debates to have an impact on the functions that the associated predicates have played in all natural languages. However, it took about another millen-

[41] See Dale Purves and R. Beau Lotto, **Why We See What we Do Redux: A Wholly Empirical Theory of Vision**, Sinauer Associates, Sunderland 2011, for many examples.

nium after Aristotle for the metaphysical implications of the logical puzzles about the properties these predicates name to be recognized — in particular the Cretan liar paradox: *All Cretans are liars* spoken by a Cretan*. True or false?* Even then, and for a long period afterwards, the puzzle did not capture the attention of logicians or metaphysicians, though philosophers like Aristotle were already troubled by the way future contingencies and vague predicates undermine the logical law of excluded middle and the principal of bivalence that seemed central to the nature of truth and falsity. It wasn't until the end of the ninetieth century that logic and metaphysics found themselves face to face with inescapable problems about the nature of truth and our knowledge of truths about truth.

The Cretan liar paradox was transformed by Frege and Russell into a fundamental challenge to our treatment of the predicate "__ is true" as the name of a property. The steps taken to construct a predicate that could confidently be said to name a property that avoid self-referential paradoxes had remarkable pay-offs in the twentieth century, among them Gödel's incompleteness theorem and Tarski's treatment of "__ is true" as the name of a metalinguistic property. But the failure to solve the problem of the Cretan liar and its more formal variants led to a significant set of challenges to the correspondence theory of truth. Pragmatic theories characterized truth as a property of the set of sentences, statements or propositions accepted at the end of (scientific) inquiry. Coherentist conceptions held that truth is a property of each of the members of the maximally consistent sentences or statements describing reality. Both of these theories convert "__ is true" into an epistemic predicate. [42] As such they may be accused simply of changing the subject from the descriptive "realist" project of identifying the property named by the truth-predicate to a recommendation of how it might hereafter be employed as an epistemic one.

Much more responsive to the problems of identifying the property named by the truth predicate have been the deflationary theories that simply deny that "__is true" names a property at all. On these theories, locutions of the form "p is true" do not attribute a property to a sentence, statement or proposition: they simply restate or reiterate the proposition. "__ is true" is a device for disquotation. Such

[42] It's worth noting that the eliminativist can provide an epistemology free from intentionality that makes it possible to take the pragmatic theory of truth as ultimately an epistemological instead of a metaphysical one. See Alexander Rosenberg, "Naturalistic Epistemology for Eliminative Materialists", *Philosophy and Phenomenological Research* 1999, Vol. 59, No. 2, pp. 335–358.

theories immediately require an account of the emergence and/or function of a predicate that is in effect redundant and adds nothing to the sentences in which it figures. One obvious response is the role of the predicate in "semantic ascent" [43] and other conveniences it offers for speakers concurring with, dissenting from and generalizing about sentences they and others speak. Deflationary theories reject not just the correspondence theory, but other "realist" claims that our employment of the truth-predicate commits us to an independent reality that thoughts are about. They do not, of course, embrace the contrary theses: that there is no independent reality or that our thoughts have intentional content. They only reject the argument that the use of the truth-predicate makes these commitments unavoidable.

Eliminativists should infer from the twentieth-century controversies surrounding the nature of truth and the predicate "_ is true" that the word does not have a sufficiently agreed-upon status to figure in a decisive argument against the coherence of their views. And they may apply this insight directly to the argument put forward within the pragmatic inconsistency objection advanced most fully by Boghossian. His argument has it that, first, eliminativists' denial that there are propositional attitudes is the assertion that mental states lack truth-conditions, and second, that thinking or saying this requires the eliminativist to hold that "There are no states with truth conditions" is true and therefore has truth-conditions — ones that bivalence does not permit to be satisfied.

Eliminativists may help themselves to disquotational theories to undermine this claim. But, Boghossian writes,

> [a] non-factualism about any subject [including of course propositional attitudes] presupposes a conception of truth richer than the deflationary: it is committed to holding that the predicate "true" stands for some sort of real, language independent property, eligibility for which will not be certified solely by the fact that a sentence is declarative and significant. Otherwise there will be no understanding its claim that a significant sentence, declarative in form, fails to possess truth conditions. [44]

[43] See Willard Van Orman Quine, **Word and Object**, MIT Press, Cambridge 1961.

[44] Boghossian, "The Status of Content…", p. 165. Boghossian's footnote to this paragraph is worth reading: "Whether truth is robust or deflationary constitutes the biggest decision a theorist of truth must make. But decide he must. It is an assumption of the present paper that the concept of truth is *univocal* as between these two conceptions, that a concurrent commitment to *both* a robust and a deflationary concept of truth would be merely to pun on the word «truth». We should not confuse the fact that it is now an open question whether truth is robust or deflationary for the claim that it

So, eliminativists cannot help themselves to deflationism to protect their denial that thoughts have truth-values from pragmatic contradiction. Accordingly, Boghossian concludes, they cannot escape the self-referential paradox involved in offering a claim, with distinct truth conditions, that there are no claims with truth conditions (i.e. with content).

Eliminativists will appreciate that Boghossian's conclusion rests on an overconfidence that correspondence or other realist theories of truth will be vindicated. The semantic, liar and self-referential paradoxes and the failure of a hundred years of attempts to solve them should undermine this confidence. There are enough grounds for disquotational theories of truth to suggest that if "__ is true" does name real property, we don't have any good idea of what that property is. And, accordingly, we should not take seriously arguments that eliminativism is incoherent based on the overconfidence of adherents of any one theory about the truth-predicate.

## 6. Conclusion: Marching Orders for Eliminativism

Eliminativists' rejection of the argument from pragmatic inconsistency shouldn't be advanced merely as a *tu quoque*. Nothing is gained merely by noting that if eliminativism is incoherent, then it is no worse off than theories of truth that are subject to the liar paradox.

And wrapping itself in the mantle of deflationary or disquotational theories of truth deprives eliminativists of obvious resources in any attempt to articulate the scientific realism they embrace in order to contrast theories in natural science with the theory of mind and folk psychology generally. [45]

Eliminativists need resources that are at least sufficient to deny scientific status to the theory of mind and other hypotheses crediting brains with propositional attitudes. And in their denial of such a status to the theory of mind, they will

can be both. There is no discernible plausibility in the suggestion that the concept of a correspondence between language and world and the concept of a language-bound operator of semantic ascent might both be versions of the same idea" (Boghossian, "The Status of Content…", p. 165 fn. 17 [emphases in the original]).

[45] Eliminativists must be scientific realists. Like instrumentalists about scientific theories, eliminativists must allow the practical utility of intentional concepts — Dennett's intentional stance. It is their realism that forbids the acceptance of hypotheses merely on the basis of their instrumental utility.

need to employ theories from the life-sciences (Darwinian theory) and the physical sciences (electrochemistry). They therefore need a basis on which to employ one set of theories in order to reject another. We have seen that they should not uncritically help themselves to notions of truth and falsity to make the invidious distinction they require between theories that are to be vindicated and ones that are to be rejected. The one-place predicates "__ is true" and "__ is false" do not appear to pick out real properties any more than do the two-place predicates "__believes that __" and "__ desires that __".

Proceeding in the way natural science suggests, eliminativists need to begin by recognizing that they face an explanatory challenge, and then respond to it in the way natural scientists do. Let us return to the earliest observation in this paper regarding the theory of mind. Its predictive weakness both in precision and range was the initial source of the eliminativist's dissatisfaction. Coupled with the ever-increasing range and precision of prediction in the physical and life sciences, this difference in predictive power calls for explanation. Churchland [46] famously extracted a relevant lesson from a similar contrast in physical science. The Ptolemaic system was employed over a millennium despite its predictive weakness owing to the absence of a better theory. It took a predictively more powerful theory to reveal that Ptolemaic theory was wrong and that its wrongness was due to the nonexistence of its explanatory variables: cycles and epicycles, deferents and equants.

If the predicates "__ is true" and "__ is false" identified real properties of sentences, statements, or propositions, then eliminativists, and everyone else for that matter, could explain the difference between successful theories and unsuccessful ones by appeal to such properties. Alas, we have seen that the truth-predicate and the falsity-predicate may not identify real properties that will explain differences in the predictive precision of theories, owing to the logical puzzles and problems that they raise.

Eliminativists, like scientists in general, should proceed by postulating, hypothesizing some real property or other, that distinguishes predictively successful theories, especially ones developed in the natural sciences, from other relatively weak, unimprovable, theories. These defective theories will typically be ones that are hard to operationalize, whose regularities have unrefinable *ceteris paribus*

[46] See Churchland, "Eliminative Materialism and the Propositional Attitudes…".

clauses, and that cannot be systematically linked to laws, theories, models and hypotheses in the natural sciences. Label the property that explains the predictive success of theories with the predicate "__ is *eurt*", and label the property that explains the predictively weak and unimproving theories "__ is *eslaf*". Now begin the search for the nature of these two properties. It is safe to assume that there is some real difference between theories that are *eurt* and those that are *eslaf*. In our search for these properties we should avoid hypotheses about them that make it impossible to employ the properties to explain the differences between predictively successful theories and predictively unsuccessful ones. And if, meanwhile, philosophers of logic and of language come up with coherent accounts of properties that the predicates "__ is eurt" and "__ is eslaf" spell backwards, well then, eliminativists will be glad to adopt these accounts in order to make better sense of the findings of neuroscience.

## Acknowledgments

For comments and criticism of previous drafts, thanks to Walter Sinnott-Armstrong, John Bickle, Dan Ross, William Ramsey, Nicholas Shea and three anonymous referees for this journal.

*Alex Rosenberg*

## References

Azizi Amir H., Wiskott Laurenz, and Cheng Sen, "A Computational Model for Preplay in the Hippocampus", *Frontiers of Computational Neuroscience* 2013, Vol. 7, article number: 161, https://doi.org/10.3389/fncom.2013.00161.

Bayne Tim and Montague Michelle (eds.), **Cognitive Phenomenology**, Oxford University Press, Oxford 2011.

Berg Nathan and Gigerenzer Gerd, "As-If Behavioral Economics: Neoclassical Economics in Disguise?", *History of Economic Ideas* 2010, Vol. 18, No. 1, pp. 133–165.

Boghossian Paul A., "The Status of Content", *Philosophical Review* 1990, Vol. 99, No. 2, pp. 157–184.

Carruthers Peter and Smith Peter K. (eds.), **Theories of Theories of Mind**, Cambridge University Press, Cambridge 1996.

Chalmers David J. (ed.), **Philosophy of Mind: Classical and Contemporary Readings**, Ox-

ford University Press, Oxford 2002.

Churchland Patricia S., **Neurophilosophy: Toward a Unified Science of the Mind/Brain**, MIT Press, Cambridge 1986.

Churchland Paul M., "Eliminative Materialism and the Propositional Attitudes", *Journal of Philosophy* 1981, Vol. 78, No. 2, pp. 67–90.

Dawkins Richard, **The Extended Phenotype: The Long Reach of the Gene**, Oxford University Press, Oxford 1982.

Dragoi George, "Internal Operations in the Hippocampus: Single Cell and Ensemble Temporal Coding", *Frontiers in Systems Neuroscience* 2013, Vol. 7, article number: 46, https://doi.org/10.3389/fnsys.2013.00046.

Dretske Fred, **Naturalizing the Mind**, *Bradford Books*, MIT Press, Cambridge 1995.

Godfrey-Smith Peter, "Review of Brian Skyrms' **Signals**", *Mind* 2012, Vol. 120, No. 480, pp. 1288–1297.

Grice Paul, "Meaning", in: Grice, **Studies in the Way of Words...**, pp. 213–223.

Grice Paul, **Studies in the Way of Words**, Harvard University Press, Cambridge 1989.

Horgan Terence E. and Tienson John L., "The Intentionality of Phenomenology and the Phenomenology of Intentionality", in: Chalmers (ed.), **Philosophy of Mind...**, pp. 520–533.

Huys Quentin J.M., Eshel Neir, O'Nions Elizabeth, Sheridan Luke, Dayan Peter, and Roiser Jonathan P., "Bonsai Trees in Your Head: How the Pavlovian System Sculpts Goal-Directed Choices by Pruning Decision Trees", *PLOS Computational Biology* 2015, Vol. 8, No. 3, e1002410.

Jacquette Dale, "Sensation and Intentionality", *Philosophical Studies* 1985, Vol. 47, No. 3, pp. 429–440.

Joffe Michael, "Mechanism in Behavioral Economics", *Journal of Economic Methodology* 2019, Vol. 26, No. 3, pp. 228–242.

Kandel Eric R., "The Molecular Biology of Memory Storage: A Dialog between Genes and Synapses", Nobel Lecture 2000, December 8, https://tiny.pl/93nvv [02.03.2022].

Kruschke John K., "ALCOVE: An Exemplar Based Connectionist Model of Category Learning", *Psychological Review* 1992, Vol. 99, No. 1, pp. 22–44.

Kubie John L. and Fox Steven E., "Do the Spatial Frequencies of Grid Cells Mold the Firing Fields of Place Cells?", *Proceedings of the National Academy of Sciences USA* 2015, Vol. 112, No. 13, pp. 3860–3861, https://doi.org/10.1073/pnas.1503155112.

Lewis David, **Convention**, Harvard University Press, Cambridge 1969.

Lewis David, "New Work for a Theory of Universals", *Australasian Journal of Philosophy* 1983, Vol. 61, No. 4, pp. 343–377.

Mante Valerio, Sussillo David, Shenoy Krishna V., and Newsome William T., "Context-Dependent Computation by Recurrent Dynamics in Prefrontal Cortex", *Nature* 2013, Vol. 503, pp. 78–84.

Moser Edvard I., "Grid Cells and the Entorhinal Map of Space", Nobel Lecture 2014, December 7, https://tiny.pl/93n6b [02.03.2022].

Moser Edvard I., Roudi Yasser, Witter Menno P., Kentros Clifford, Bonhoeffer Tobias, and Moser May-Britt, "Grid Cells and Cortical Representation", *Nature Reviews Neuroscience* 2014, Vol. 15, pp. 466–481.

Moser May-Britt, "Grid Cells, Place Cells, and Memory", Nobel Lecture 2014, December 7, https://tiny.pl/93nvq [02.03.2022].

Nichols Shaun, Stich Stephen, Leslie Alan, and Klein David, "Varieties of Off-Line Stimulation", in: Carruthers and Smith (eds.), **Theories of Theories of Mind...**, pp. 39–74.

O'Keefe John, "Spatial Cells in the Hippocampal Formation", Nobel Lecture 2014, December 7, https://tiny.pl/93nvm [02.03.2022].

O'Keefe John and Dostrovsky Jonathan, "The Hippocampus as a Spatial Map: Preliminary Evidence from Unit Activity in the Freely-Moving Rat", *Brain Research* 1971, Vol. 34, No. 1, pp. 171–175.

Paul L.A., "Categorical Priority and Categorical Collapse", *Proceedings of the Aristotelian Society* 2013, Vol. 87, Supplementary Volumes, pp. 89–113.

Purves Dale and Lotto R. Beau, **Why We See What we Do Redux: A Wholly Empirical Theory of Vision**, Sinauer Associates, Sunderland 2011.

Quine Willard Van Orman, **Word and Object**, MIT Press, Cambridge 1961.

Rosenberg Alexander, "Naturalistic Epistemology for Eliminative Materialists", *Philosophy and Phenomenological Research* 1999, Vol. 59, No. 2, pp. 335–358.

Rushworth Matthew F.S., Noonan Maryann P., Boorman Erie D., Walton Mark E., and Behrens Timothy E., "Frontal Cortex and Reward Guided Learning and Decision Making", *Neuron* 2009, Vol. 70, No. 6, pp. 1054–1069.

Searle John R., **Intentionality: An Essay in the Philosophy of Mind**, Cambridge University Press, Cambridge 1983.

Searle John R., "Minds, Brains and Programs", *The Behavioral and Brain Sciences* 1980, Vol. 3, No. 3, pp. 417–424.

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

Skyrms Brian, **Signals: Evolution, Learning and Information**, Oxford University Press, New York 2010.

Stich Stephen, "Do True Believers Exist? A Reply to Andy Clark", *Aristotelian Society Supplement* 1991, Vol. 65, pp. 229–244.

Stich Stephen, **From Folk Psychology to Cognitive Science**, MIT Press, Cambridge 1983.

Strawson Galen, "Cognitive Phenomenology: Real Life", in: Bayne and Montague (eds.), **Cognitive Phenomenology...**, pp. 285–325.

Szabo Gendler Tamar and Hawthorne John (eds.), **Perceptual Experience**, Oxford University Press, Oxford 2006.

Unger Peter, "There Are No Ordinary Things", *Synthese* 1979, Vol. 41, No. 2, pp. 117–154.

Van Essen David C. and Gallant Jack L., "Neural Mechanisms of Form and Motion Processing in the Primate Visual System", *Neuron* 1994, Vol. 13, No. 1, pp. 1–10.

Van Inwagen Peter, **Material Beings**, Cornell University Press, Ithaca 1990.

Yu Jai Y. and Frank Loren M., "Hippocampal-Cortical Interaction in Decision Making", *Neurobiology of Learning and Memory* 2015, Vol. 117, pp. 34–41, https://tiny.pl/93nbq [02.03.2022].

Yu Jai Y., Kay Kenneth, Liu Daniel F., Grossrubatscher Irene, Loback Adrianna, Sosa Marielena, Chung Jason E., Karlsson Mattias P., Larkin Margaret C., and Frank Loren M., "Distinct Hippocampal-Cortical Memory Representations for Experiences Associated with Movement *versus* Immobility", *eLife* 2017, Vol. 6, e27621, https://doi.org/10.7554/eLife.27621.