

# Naturalizing Metaphysics with the Help of Cognitive Science

Alvin I. Goldman

Forthcoming in Karen Bennett and Dean W. Zimmerman (eds.),  
Oxford Studies in Metaphysics (vol. 9, 2014)

## 1. Introduction

This paper advances a thesis in the methodology of metaphysics. It argues that empirical findings in cognitive science can play a significant evidential role in an optimal methodology for metaphysics. Metaphysicians therefore have an epistemic interest in being attuned to appropriate evidence from cognitive science. This may even rise to the level of epistemic obligation, because epistemic inquiry in general -- of which the pursuit of metaphysical truth is an instance -- requires responsible inquirers to heed highly relevant and available evidence.<sup>1</sup> Acquiring cognitive scientific evidence can (sometimes) precipitate rational changes in credence functions that metaphysicians assign to competing theories. However, I do not mean to propose any utterly radical metaphysical methodology or any wholesale replacement of traditional methods. The proposal is conceived of as a supplement to traditional methods rather than a replacement of them.

What is proposed, then, is a partial "naturalization" of the methodology of metaphysics. I distinguish between three main kinds of contrasting methodologies that philosopher-metaphysicians can adopt. The first methodology is to follow the traditional conception of metaphysics as a purely a priori, armchair enterprise. It would find no place for any scientific input into metaphysical deliberation. A second methodology would welcome contributions to metaphysics from physics and other (non-biological, non-psychological) physical sciences; but not from cognitive science. This form of limited naturalization might be defended as follows.

Cognitive science is uncontroversially relevant to the metaphysics of mind. Questions about consciousness, the nature and ontological status of phenomenology, the relation between mental states and brain states -- all of these questions invite appropriate

inputs from cognitive science. When the question is raised, however, about how to do metaphysics in general -- precisely the question being raised here -- why would cognitive science enter the picture? What kinds of information could cognitive science give us about the "external world"? The customary answer would be: none. So the second methodological position would also exclude cognitive science from playing a significant role.

The second position is compatible, however, with an extensive amount of "partial" naturalization of metaphysics, an approach that has made substantial gains in recent metaphysical practice. Appeals to physics have become fairly routine, or at least not uncommon, in the literature. At the radical end of the spectrum is James Ladyman and Don Ross's book Every Thing Must Go: Metaphysics Naturalized (2007). It proposes to infuse metaphysics with physics and essentially abandon everything else -- at least the usual types of conclusions and methodologies used by traditional metaphysicians. A less extreme program for infusing physics into metaphysics is Tim Maudlin's The Metaphysics within Physics (2007), which describes itself as containing "the outlines of an ontology based on physics" (2007: 3). Yet another example in this category is Jonathan Schaffer's (2010a, b) use of physics to defend a very traditional metaphysical thesis -- i.e., monism -- by appeal to a detailed account of quantum entanglement. Schaffer traces the roots of monism to Parmenides, Plato, Plotinus, Spinoza, Hegel, and Bradley.

With a few exceptions, there is very little in the literature that supports a general role for cognitive science within metaphysics, including a role in the methodology of metaphysics.<sup>2</sup> More precisely, few metaphysicians who pursue a realist as opposed to a conceptualist program of metaphysics display much interest in cognitive science. By "conceptualist" metaphysics I mean the project of delineating the ontology (and ideology) characteristic of naïve human thought, i.e. folk ontology. There is a growing literature in this conceptualist vein, concerned with the naïve understanding of metaphysical concepts and phenomena. This approach, which saliently includes work by experimental philosophers, welcomes inputs from cognitive science. (Experimental philosophy might itself be classed as a sector of cognitive science.) But this work is not realist metaphysics. It does not display the dominant mission of mainstream metaphysics,

which is to characterize the correct or true metaphysics of the world -- as opposed to what is commonsensically thought or said about these matters. Metaphysicians of the realist stripe might fear that cognitive science material can only be informative about modes of thought, whereas proper metaphysics aims to understand, not thought, but the world's objective character. So they may choose to occupy the second methodological position explained above, which would still exclude cognitive science from the methodology of metaphysics generally. I wish to embrace the realist mission of metaphysics.<sup>3</sup>

Nonetheless, I shall argue, even a metaphysics that sails under a realist banner is best conducted with the help of cognitive scientific inputs. This is a third kind of methodology, which I shall defend here.

As mentioned in footnote 2, some other metaphysicians have recently made similar pitches for selected areas of metaphysics. My aim is broader: to offer a general template (or two) for metaphysical methodology under which cognitive scientific considerations might become routine or commonplace factors in realist metaphysics, not just isolated or occasional factors, as matters currently stand. It is these proposed templates that comprise the chief innovations of the present paper, not the particular applications of them that are presented.

I start with a brief heuristic rationale for the general idea. Like other types of philosophers, metaphysicians appeal extensively to intuition, experience, and commonsense belief to guide their path in metaphysical theorizing. Most cognitive scientists, however, contend that intuitions are massively influenced by our cognitive system -- or "cognitive engine," as I shall call it. They are the products of complex computational operations, or neural circuits, which have formed over eons of evolutionary time. The cognitive outputs are rarely if ever simple read-outs of sensory inputs. Instead, they tend to be artifacts of "biases" or "constraints" wired into our cognitive equipment. In seeking to characterize the world itself, therefore, we had better not ignore the features of the complex systems that mold and shape our perceptual and cogitative experience. This means paying attention to the deliverances of cognitive science, our best if not exclusive source of information about the underlying systems.

To fix ideas, consider a familiar illusion from perceptual science, the Kanisza illusion, found in many elementary perception textbooks. (See Figure 1)

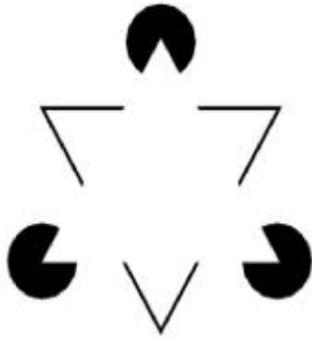


Figure 1

This visual display is compellingly seen as containing a vivid white triangle with sharp contours in the foreground. Physically speaking, however, there are no such contours on the page (or screen). In fact the triangle's contours are perfectly homogeneous in color with their immediate white surrounds. The observed sides of the triangle are merely apparent; not genuine marks on the page (or screen). The visual experience of them is created or produced (automatically) by the visual system. Other visual illusions, of course, abound. They testify to the fact that our cognitive engine plays a significant role in shaping or "filling in" what is experienced. So we should not blithely assume that experience reliably signals, or indicates, features of the external world. A more complex approach is needed when we reflectively consider the nature of external reality.

Provision must be made for the fact that what appears spontaneously to be features of the external world (punkt) may be a highly constructed product of an autonomous cognitive engine, the silent operations of which are hidden from perceptual or introspective view. This is not a new story, but it is still a good -- and important -- story. It is a significant reason why cognitive science should not be neglected.

This story can be given a slightly more formal statement in terms of a general functional relationship between genuine metaphysical reality, on the one hand, and our pre-philosophical intuitions or apprehensions about objects, properties, and relations that populate reality. Let "CSE" (commonsense experience) refer to the vast set of experiences and representations people have concerning reality, i.e., perceptions,

intuitions, and judgments concerning the world (expressed in both natural language and the language of thought, if such there be). Examples include experience (as) of the passage of time, apprehension and/or judgment of object persistence over time, and judgments about the number of clouds currently in the sky. Let "R" refer to reality as a whole, or the class of objects, properties, and relations that (genuinely) populate reality. As previously indicated, a fundamental assumption of cognitive science is that commonsense experience and representations of reality are a function not only of reality itself but of the specific kind of cognitive engine (COGEN) we carry with us in all of our interactions with the world. As a result, relations between reality and our perceptual, intuitional, and cognitive acts cannot be adequately captured by a simple functional schema like (1):

$$(1) \text{ CSE} = f(R).$$

It should instead assume a form more like (2):

$$(2) \text{ CSE} = f(R, \text{COGEN}).^4$$

Arguably, formula (2) should also be expanded in the direction of (3):

$$(3) \text{ CSE} = f(R, \text{COGEN}, \text{CUL})$$

in which CUL is the culture in which the cognizer is embedded. For present purposes, however, formula (2) is instructive enough. If commonsense experience were a function of reality alone (i.e., the "rest" of the reality, minus the brain), the metaphysical investigator might hope that each slice or pattern of experience is a determinate function of (external) reality plus our interactions with it -- as suggested by formula (1). If  $f(R)$  could be specified, an inquiring metaphysician might be able to make an inverse inference from CSE to R. Even this would be problematic, of course, since there may be no determinate inverse function. The same pattern of experience might be compatible with alternative forms of reality. The situation is more complicated yet, however, if

formula (2) rather than formula (1) represents the state of play. If two very different kinds of factors are responsible for a given pattern of experience -- both the nature of reality and the character of our perceptual and cognitive equipment -- then problems ramify for a seeker of metaphysical reality. If our aim is to infer the nature of reality yet our impressions of reality are initially shaped by our cognitive engine (of which we can never divest ourselves), an appropriate epistemic route to the goal of getting a good grip of reality is to pay careful attention to the "silent" role that the cognitive engine plays in generating cognitive outputs. If this engine uses operating principles that help mold the cognitive outputs, this fact should not be neglected in our inquiry. It might be too easy to misattribute (apparent) properties generated by the engine to reality itself. Whatever evidence can be uncovered, then, about the mind's operating principles should be incorporated into a search for the properties of reality.

An analogy from a very different domain may convey the spirit of the perspective I mean to encourage. If a voter ponders the future actions and practices of a political candidate, what kinds of evidence should the voter try to track down and deploy? Only a deplorably naïve voter would rely exclusively and uncritically on a candidate's public statements and promises. An informed voter recognizes the lengths to which campaign organizations go to package their candidate for the electorate's consumption. A voter must look beyond such packaging to infer what lies behind it. The present suggestion is that a savvy metaphysician should do no less. What lies "behind" commonsense intuitions, experiences, and judgments (which metaphysicians use as *prima facie* guides to the nature of reality) should be probed, including what can be gleaned from the scientific study of the cognitive engine.

This is roughly the story advanced by Kant (on a familiar rendition, at any rate), who also promoted the "critical" study of metaphysics. In our own period, the credo of cognitive science was shaped by the idea that outputs of the cognitive engine are not simple products of an external "stimulus." Focusing on knowledge of grammar, Chomsky (1980) argued for the "poverty of the stimulus". On this view the mind has an innate schematism that places substantive constraints on what kinds of cognitions are generated. This innate endowment accounts for the fact that young children converge on the same grammar, despite encountering a very small amount of data. Not all

practitioners of cognitive science endorse Chomsky's precise version of the thesis. Nonetheless, cognitive science broadly embraces an attenuated version of it, to the extent of highlighting the powerful impact of the mind's autonomous structure in generating surface-level cognition. Human cognition is constrained and shaped by a specific set of organs, which channel and give rise to distinctively human kinds of cognitive outputs. This idea should be taken to heart by inquiring metaphysicians no less than they take to heart the fundamentals of modern physics.

## 2. A Bayesian Approach to Revising Metaphysical Credences

How does this perspective translate into methodological advice for metaphysicians? Disputes in metaphysics are often presented as choices between how to interpret experience and discourse in particular domains. What properties, relations, facts, truths, or states of affairs obtain, hold, or are exemplified in each domain? The options on offer for metaphysical choice are frequently couched in the language of "realism" versus "anti-realism" (within a specified domain). The realism in question is a somewhat different sense than the one referred to in section 1. It is commonly understood as a view that upholds the existence of facts, truth-makers, or mind-independent properties, and which correspond fairly directly to our naive conception of the domain. In contrast to realism (in this sense), an assortment of anti-realisms are on offer for many metaphysical domains. Terminologies for the variety of anti-realisms vary from domain to domain and from author to author. Popular terminologies for anti-realisms include eliminativism, nihilism, dispositionalism (or response-dependence), expressivism, fictionalism, projectivism, constructivism, error theories, and so on. I won't try to give precise definitions of each of these terms. In any case, the main purpose here is not to decide which of these alternatives is correct in each domain, but what methodology to use in deciding which is correct. More specifically, what kinds of evidence should metaphysicians consult to make optimal choices? More specifically yet, the central question is whether evidence from cognitive science can be particularly helpful or even essential. If so, what kinds of findings from cognitive science might be evidentially relevant to metaphysical choices, and exactly how are they relevant?

In many domains of metaphysical dispute, defenders of realism are confronted with opponents who seek to debunk realism.<sup>5</sup> In meta-ethics realism about moral value is contrasted with various kinds of anti-realism. Philosophers who critique realism -- from the perspective of evolutionary theory, for example -- often call themselves debunkers. Sharon Street (2006) and Richard Joyce (2006) are debunkers in this sense. They defend anti-realism about moral value, the view that there are no antecedent, mind-independent evaluative facts about which things are reasons for doing things. According to such anti-realists, facts about moral reasons consist in, or rest upon, people's attitudes, attitudes that are the products of evolution rather than (for example) the detection of things in mind-independent reality.

How might such "debunking" take place? Roger White (2010) notes that many debunking attempts are based on alternative causal explanations of the (realist) beliefs in question. Street and Joyce defend evolution-based anti-realism as an alternative to a realist, detectivist account of moral value. White is skeptical about such debunking attempts.<sup>6</sup> But despite his skepticism, he introduces a simple but instructive example, which I shall modify for purposes of my own argument. Here is White's example:

That the gas gauge reads Full is evidence that the tank is full. But now I learn that the gauge is stuck and would indicate Full whether the tank is full or not. In this way my belief formation was not sensitive to the truth as the gas gauge on which I based my belief was not sensitive to the contents of the tank. This information undercuts the evidential support of the gauge reading. And if my justification for believing the tank is full crucially depended on this evidence then I am no longer justified in supposing that the tank is full.... (2010: 585).

White describes the case in terms of categorical belief rather than credences and categorical support or defeat rather than probabilities. But the story could easily be re-analyzed in Bayesian terms. I want to explore a quasi-Bayesian approach in this section, so let us see how such an approach would work here.

Re-describe the gasoline example as featuring two competing hypotheses: the gas tank is full (H) and the gas tank is less than full ( $\sim$ H). Assume that your friend has just

lent you his car on an emergency basis, and you have no idea about its fuel situation. For simplicity, then, assume that the priors you assign to H and  $\sim H$  are equal: both are .50. Upon entering the car, you consult the gas gauge and see that it says "Full", and take that as an item of evidence. You assess the crucial likelihoods (conditional probabilities) as follows:

(L<sub>1</sub>) The probability that the gas gauge would read "Full" (F) conditional on the gas tank being full ( $P_H(F) = 0.9$ )

(L<sub>2</sub>) The probability that the gas gauge would read "Full" conditional on the gas tank being less than full ( $P_{\sim H}(F) = 0.1$ ).

Given these priors and likelihoods, the posterior probabilities you will assign to H and  $\sim H$  (given that the gauge reads "Full") are as follows:  $P_F(H) = 0.90$  and  $P_F(\sim H) = 0.10$ . In other words, you regard it as highly probable that the tank is full.

Now you learn that the gas gauge is stuck. The first epistemic effect of this new evidence, we may suppose, is that you revise the likelihoods you associate with the gas gauge reading full.<sup>7</sup> Exactly how such revisions should be made is not obvious in the absence of additional information, but minimally both likelihoods should be modified, presumably in the direction of more intermediate values. Assume you revise both L<sub>1</sub> and L<sub>2</sub> so that L<sub>1</sub> = L<sub>2</sub> = .50. In other words:

(L<sub>1</sub>')  $P_H(F) = .50$ .

(L<sub>2</sub>')  $P_{\sim H}(F) = .50$ .

To see how one's credence function should be revised in light of the revised pair of likelihoods, we simply plug the new values into a standard form of Bayes' Theorem. This form and the appropriate values are given here:

$$P_F(H) = \frac{P(H) P_H(F)}{P(H) P_H(F) + P(\sim H) P_{\sim H}(F)} = \frac{(.5) (.5)}{(.5)(.5) + (.5)(.5)} = \frac{.25}{.50} = .50.$$

Thus, having learned about the stuck gas gauge, your new posterior probabilities are  $H = \sim H = .50$ . Under the assumed credal assignments, the epistemic impact of acquiring the new evidence about the stuck gas gauge is considerable.

This analytical template is what I wish to propose for applying cognitive science to metaphysical disputes (at least disputes of the realism/anti-realism variety). In the foregoing illustration, of course, the new evidence has nothing to do with cognitive science. But in the examples that follow, the evidence of interest will be products of cognitive science. Again, in the foregoing example, the result of acquiring the new evidence pushes the rational cognizer toward agnosticism about the target hypothesis. I do not mean to imply that cognitive scientific findings will always breed agnosticism (equal credences vis-à-vis competing theories). Nor do I mean to suggest that such findings would be decisive for settling the metaphysical dispute. I only suggest that suitable cognitive scientific findings can force a rational metaphysician to adjust his/her credences in light of those findings. Such findings will be relevant evidence for credal adjustments vis-a-vis the metaphysical hypotheses in question. This is one template for the central thesis I mean to defend in this paper.

I shall next apply this template to four problems in metaphysics. To effect these applications, several kinds of substitutions need to be made. Instead of the gas-tank hypotheses  $H$  and  $\sim H$ , we would need rival hypotheses of a metaphysical kind. I am specifically interested in pairs of metaphysical hypotheses, or theories, that compete as accounts of designated metaphysical phenomena. One would be a realist theory and the other an anti-realist theory. The metaphysician's initial evidence bearing on the theories would be a set of ordinary experiences, intuitions, or beliefs about the domain that each theory tries to accommodate in its own way. A basic assumption here is that such experiences, intuitions, and beliefs are examples of evidence that metaphysicians (legitimately) use when weighing competing metaphysical theories. Finally, I shall identify cases in which cognitive science has (already) found evidence that (by my lights) should influence the credences or credence functions assigned to the likelihoods that link theories and experiential evidence, on the model of the stuck-gauge evidence in White's example. In principle, my points could be made with non-standard metaphysical theories of my own devising and purely invented cognitive scientific findings. But, wherever

possible, I shall work with actual metaphysical debates between realist and anti-realist alternatives and with actual cognitive-scientific findings. Of course, such empirical findings are always open to debate and interpretation. But I seek to use examples with pretty legitimate scientific standing.

### 3. First Illustration: The Passage of Time

The metaphysics of time, since McTaggart (1908), has been dominated by two competing theories: the A-theory and the B-theory. The A-theory is commonly taken to imply that tense is real, that there is something privileged about the present, and (often) that time flows or passes, a notion entirely absent from the B-theory. One lively example of a realism/anti-realism debate is precisely the debate over the reality or unreality of temporal passage. Arguments on behalf of this conclusion vary, but many philosophers contend that we perceive such passage in our perception of change. Thus, Richard Schuster (1986) writes:

Let me begin this inquiry with the simple but fundamental fact that the flow of time, or passage, as it is known, is given in experience, that it is as indubitable an aspect of our perception of the world as the sights and sounds that come in upon us, even though it is not the peculiar property of a special sense. (1986: 695)

L. A. Paul (2010) introduces her treatment of the topic as follows:

We all know what it is like to have [passage-of-time experiences]. Reflection on the qualitative character of such an experience suggests that events occurring now have a characteristic property of *nowness*, responsible for a certain special "feel," and that events pass from the future to the present and then into the past. The question that I want to explore is whether we should take this suggestion to support an *antireductionist* ontology of time, that is, whether we should take it to support an ontology that includes a primitive, monadic property of *nowness* responsible for the special feel of events in the present, and a relation of *passage*

that events instantiate in virtue of literally passing from the future to the present and then into the past. (2010: p. 333)

What Paul calls "antireductionism" (with respect to time) is a view that takes both passage and nowness to be real; it is obviously the realist position in the debate. What she calls "reductionism" is the antirealist position. It claims that there is no such property as passage (or nowness; but we'll focus on passage). According to antirealism, we just inaccurately experience the world as having this property. What evidence can the competing sides invoke to support their respective cases?

Antireductionists rely on the force of the experiences themselves. There must be some sort of physical flow, they say, that we detect via our experience as of passage. The problem for reductionists, says Paul, is to provide an account of why (or how) we have such temporal experience despite there being no ontologically real passage. "It is absolutely essential for reductionists to be able to provide an alternative, reasonable explanation of why we have temporal experiences as of nowness and passage." (2010: 339). She proceeds to offer cognitive science-inspired explanations of how we might have such experiences despite there being no such things as nowness or passage.<sup>8</sup>

She first appeals to the familiar phenomenon (within cognitive science) of apparent motion. Blinking dots in rapid succession on opposite sides of a screen give rise to an illusion of a dot moving back and forth across the screen. And in the cutaneous rabbit experiment a series of appropriately spaced taps (e.g., taps at the wrist, near the elbow, and the upper arm area) produces an illusion of an object moving continuously along one's arm. The "color phi" experiment reports an analogous illusion. Extrapolating from these cases, she suggests, the reductionist can argue that our illusion-creating brain is quite capable of working with a static universe to produce experiences that suggest passage and flow though in reality there are no such phenomena.

Similarly, another opponent of "real" temporal flow, Robin Le Poidevin (2007) appeals to the "flash-lag" experiment to explain a nonveridical experience of temporal flow.

Subjects are presented with a small dot moving across a screen. At some point during the dot's transit, another dot appears briefly directly above or below

the other dot: this is the 'flash'. A significant number of subjects report that they saw the moving dot ahead of the flashed dot, in the direction of movement. The flash appears to lag behind (McKay 1958). But it does not really do so, so why the illusion that it does?

[I]t takes time for the brain to register what is going on in the outside world. When information is registered about changing objects, and especially about rapidly moving objects, it is already slightly out of date. So we have evolved to compensate for this lag. The brain makes an adjustment to the information it has received about the position of a moving object, and makes a projection based on information concerning the object's velocity and direction. When we see the object, we see it, not in the position it was in when light from it hit the retina, but in the position the brain estimates that the object must be in by the time that information is registered. What we see, once more, is a projection, based on a prediction.... Contrasted with this 'predictive' model is the 'postdictive' model... As with the predictive model, however, the suggestion is that the brain is imposing an interpretation on the data. (Le Poidevin 2007: 94)

Thus, both Paul and Le Poidevin offer explanatory stories to defend anti-realism. I completely accept the relevance of such explanatory factors. But as an analytical device I instead propose Bayesianism (as contrasted with explanationism) as a more illuminating and perspicuous account of how metaphysicians should proceed, epistemologically, in the realism/anti-realism debate. What's the connection between the two?

Suppose that anti-realism is true: passage is not ontologically real. If under this supposition there would be no psychological or brain-based explanation of our having experiences as of passage (despite its ontological unreality), it would be entirely reasonable for metaphysicians to assign a very low number to the conditional probability (likelihood) of our undergoing passage experiences if anti-realism were true. On the other hand, if there were a plausible potential explanation of why or how we would undergo passage experiences (even) if antirealism were true, a metaphysician would then have good reason to assign a much higher number to the same conditional probability. Such a boost in conditional probability assignment could make a big difference in the

posterior probabilities that reasonable metaphysicians would assign to the realist and anti-realist hypotheses respectively. Through the demonstration of illusions such as the Kanisza illusion (section 1), cognitive science has shown that the human brain is not only capable of, but even fairly prone to, create non-veridical contours and other non-veridical effects. Thus, explanations of the occurrence of temporal passage experiences despite their non-veridicality is not at all far-fetched.

In short, by taking relevant cognitive science into account, a reasonable metaphysician would substantially revise the likelihood of the occurrence of passage experiences conditional on anti-realism. Letting  $E$  = experiences of passage,  $R$  = ontological realism about passage, and  $AR$  = ontological anti-realism about passage, we might expect the indicated types of evidence from cognitive science to boost the likelihood of  $E$  given  $AR$  from an initial value of, say, .001 (choosing a number out of a hat) to .40 (choosing another number out of the hat). This would make a very significant difference to the resulting posterior probability that a metaphysician informed by cognitive science (and utilizing Bayesian principles) would attach to anti-realism about passage.

Cognitive scientists would not be surprised by this sort of metaphysical move (although they might not like to describe it in the language of "metaphysics"). They routinely view the constructive activity of the brain as a massive factor in determining human experience. So, whether they would agree in detail on this particular example, the general drift of the analysis is completely congenial to their general perspective.

#### 4. Second Illustration: Moral Value

Sharon Street (2006) formulates the defining claim of value realism as the claim that there are evaluative facts or truths that hold independently of all our evaluative attitudes, whereas anti-realism denies that there are such truths. However, as Street points out, there is a basic set of evaluative judgments that human beings tend to embrace across time and culture. These include the following judgments about reasons, which I shall call, collectively, EJ (for "evaluative judgments").

- The fact that something would promote one's survival is a reason in favor of it.
- The fact that something would promote the interests of a family member is a reason to do it.
- We have greater obligations to help our own children than we do to help complete strangers.
- The fact that someone has treated one well is a reason to treat that person well in return.
- The fact that someone is altruistic is a reason to admire, praise, and reward him or her. (Street 2006: 115)

Can a metaphysician (or meta-ethicist) choose between realism and anti-realism by reference to the likelihood of the widespread embrace of these value judgments given realism versus the likelihood of their widespread embrace given anti-realism? What is the likelihood that EJ would be widely embraced if there were independent moral truths and what is the likelihood of EJ being widely embraced if there were no such independent moral truths? How, in other words, should the following likelihoods be evaluated?

$$(1) P_R(EJ) = ?$$

$$(2) P_{AR}(EJ) = ?$$

Finally, can cognitive science present any helpful evidence in these matters?

Realists would say that these judgments could easily be widely accepted if realism were correct -- that is, if there were independent moral truths -- because of people's ability to "track" them. But could people track such truths? They might be able to do so if they had an (accurate) moral sense; but do they? Cognitive science can weigh in on this question. It certainly studies perceptual systems (also called "senses") and sometimes discovers new ones. Comparative cognitive science discovered echolocation in bats. Human cognitive neuroscience has discovered many types of interoceptive senses in the human brain, i.e., brain mechanisms for monitoring the positions and conditions of one's own bodily organs (see Craig, 2002; de Vignemont, 2011; Goldman, 2012).<sup>9</sup> There is no comparable discovery of any moral sense, whether an exteroceptive sense or an interoceptive sense.<sup>10</sup> This might be interpreted as evidence against the

existence of any sense for tracking moral truths, which presumably militates against a high likelihood, under value realism, of broad convergence on moral judgments like EJ.

Is it any more likely that such a broad convergence on moral judgments would be made under the alternative hypothesis, i.e., the non-existence of independent moral truths? Street argues that such a broad convergence -- especially convergence on the reasons listed in EJ -- is substantially likely under anti-realism because evolution provides a very plausible avenue to this convergence via natural selection. Natural selection would select for mechanisms that tend to get organisms to act, feel, and believe in reproductively advantageous ways. Pro-attitudes toward the judgments in EJ would have precisely this property, because acting in the manners described by these judgments would have clear survival and reproductive advantages. Natural selection, she points out, often selects for mechanisms that do just this, as in the case of the automatic reflex response that causes the hand to withdraw from a hot surface. She also points to "the striking continuity that we observe between our evaluative judgments and the more basic evaluative tendencies of other animals, especially those most closely related to us....[A]t some basic motivational level, chimpanzees seem to experience the fact that another chimpanzee has helped them, whether by sharing food, grooming them, or supporting their position within the group hierarchy, as 'counting in favor of' assisting that other individual in similar ways." (2006: 117). Since evolution seems to have favored our nearest animal relatives having similar traits, it would be "easy" for evolution to produce similar traits in a nearby species such as ours. For anybody not previously apprised of such facts, this is additional evidence that should boost the likelihood that EJ would be favored attitudes among humans as well. If we count evolutionary science as part of cognitive science (and the neuroscience branch of cognitive science certainly studies brain evolution), this would be another case in which evidence emerging from cognitive science is relevant to an epistemic choice between metaphysical alternatives.

Summing up, evidence from cognitive science can weigh in on both likelihoods (the one pertaining to realism and the one pertaining to anti-realism) in a substantial way. It can raise doubts about the existence of a moral sense, making it harder for a single body of moral judgments to become entrenched across time and culture. And it can make it clear how natural selection could promote the result that a single body of moral

judgments, like EJ, would become entrenched across time and culture. Bayesian inference applied to changed likelihoods would give rise to changed posterior probabilities vis-à-vis the core hypotheses in question: value realism and value anti-realism.

##### 5. Third Illustration: Natural Kinds

An extremely influential idea in contemporary metaphysics, dating to the 1970s, is the idea that natural kinds have essences that are discovered by science and that such essences determine the extensions of our natural kind terms and concepts. Certain metaphysical necessities have also been inferred from this view. For example, if it is the essence of water to have the chemical composition H<sub>2</sub>O, then it is necessarily the case that every sample of water has the chemical structure H<sub>2</sub>O. Philosophical acceptance of kind essentialism also brought in its train (or was accompanied by) acceptance of certain de re necessities such as the necessity of origin. If woman Y was in fact my biological mother, then necessarily Y was my mother. No other person could have been my mother. All this is the product of work by Hilary Putnam (1975) and Saul Kripke (1980).

How could cognitive science challenge such metaphysical theses? Sarah-Jane Leslie (2013) mounts such a challenge as follows.<sup>11</sup> The rationale for the essentialist theory of natural kinds is, fundamentally, that people find it intuitively compelling to understand objects in the world in terms of this theory. This is part of their psychology, a thesis that Leslie labels quintessentialism (= psychological essentialism). Moreover, these intuitions are the outgrowth of an implicit set of beliefs, emerging early in preschooler development. In other words, Leslie points to an extensive body of literature by developmental psychologists according to which even young children are (psychological) essentialists. Their psychologies lead them to believe in essences for a wide variety of things they encounter. An essence is understood as something like a hidden inner structure. (Important contributions and reviews to this literature include Keil, 1989, and Gelman, 2003). Leslie provides illustrations of the supporting data for the presence of this idea in early childhood. From the philosophical perspective, her principal critical point is that the existence of psychological essentialism as a cognitive trait provides an alternative explanation of people's intuitions about kinds and essences.

As Leslie puts it, "There are many explanations of why we may be fundamentally disposed to see the world in a particular way, only one of which is that metaphysically or scientifically speaking, the world actually is that way" (2013: 108; emphasis in the original). In other words, it is questionable whether the realist postulation of real-world essence is the best explanation of our essentialist intuitions. This is fully parallel to the issues raised in the first two examples of what is the (best) explanation of intuitions or attitudes that are admittedly highly prevalent or widespread. In effect, Leslie is advancing an anti-realist explanation of people's essentialist intuitions. Their prevalence is perfectly well explained by the independent existence of a certain (possibly innate) psychological trait; we don't need to postulate essentialist facts about physical objects.

Leslie does not rest content, however, with presenting the (anti-realist's) purely cognitivist explanation of essentialist intuitions. She argues independently, and at length, that physical science fails to support the essentialist views that Putnam and Kripke advance. For example, Putnam and Kripke advance the view that the necessary and sufficient condition of belonging to a natural kind is given by a hidden underlying structure found in its members. In the case of water the relevant structure (essence) is H<sub>2</sub>O and in biological species the relevant structure is the genetic code. But this is false, Leslie argues, both with respect to chemical kinds and biological kinds. With respect to biological species, a member of one species may have more genetically in common with a member of another species than with a member of its own species. And with respect to chemistry, matters are much more complicated than the suggested essence acknowledges. She then concludes as follows:

We can now pose the question: does science actually deliver such a privileged same substance relation, or do we simply have once again the quintessentialist intuition that science must do so? ... I shall argue that the relevant sciences deliver no such privileged same substance relation. (2013: 143)

Clearly, Leslie's argument is not confined to showing that cognitive science delivers relevant evidence. She also contends that physical science has a lot to offer. This part of her thesis, however, is simply independent of the part pursued here. It obviously does not

challenge, but simply complements, the evidential relevance of the cognitive science, which she also supports for her selected domain.

Nonetheless, let us see how Leslie's materials both fit within the Bayesian framework used here. As with the first two illustrations, Leslie argues for the relevance of cognitive scientific evidence to the competing metaphysical hypotheses. If we focus only on the contribution of cognitive science to her analysis, it can play the same role in a Bayesian analysis as it did in our earlier examples. There is no analogue of the deployment of physical science evidence in the previous examples, but that constitutes no important disanalogy for present purposes. After all, it is not my purpose to deny an evidential role for physical science in these metaphysical controversies.

How would a Bayesian analysis proceed, then, for Leslie's case? The cognitive science portion of her evidence contributes to the probability of realism about physical essentialism by addressing the conditional probability that people would have essentialist intuitions on the assumption that essentialist realism is false. The idea is that given the fact of psychological essentialism -- a built-in tendency to expect hidden essences -- it would still be highly probable that children and grown-ups would have essentialist intuitions even if the physical world did not rampantly feature essences. This evidence should lead a metaphysician to raise her credence in the likelihood of the widespread incidence of essentialist intuitions (even) given the anti-realist hypothesis, and this change in likelihood will lower the posterior probability of essentialist realism.

Let us put some numbers in place to illustrate this point. Let  $E$  = the regular occurrence of essentialist intuitions. Let  $R$  = essentialist realism, the hypothesis that biological and chemical kinds (for example) actually have hidden-structural essences. Finally, let  $AR$  = essentialist anti-realism, the hypothesis that biological and chemical kinds do not reveal hidden structural essences. Before Leslie brings psychological essentialism into play, how would a reasonable metaphysician evaluate the following two conditional probabilities (or likelihoods): the probability of  $E$  conditional on  $R$  and the probability of  $E$  conditional on  $AR$ ? That is, what values might plausibly be assigned to  $P_R(E)$  and to  $P_{AR}(E)$ ? The former will presumably be relatively high, at least assuming that people are scientifically pretty competent. They will acquire essentialist intuitions if the world actually has hidden essences that can be uncovered. Grown-ups, at least,

should acquire such intuitions. But the value of  $P_{AR}(E)$  should be relatively low. If the real-world has no hidden essences, why would people have essentialist intuitions? Enter now the fact that people (including pre-schoolers) have the cognitive trait of quintessentialism, which is acquired by normal maturation no matter what science is formally introduced in classrooms. Given this fact, which we assume to be contributed by cognitive science,  $P_{AR}(E)$  must be adjusted upward.

Before we get the cognitive scientific evidence, it is plausible to substitute the following numbers into the Bayesian formula:

$$P_E(R) = \frac{P(R) P_R(E)}{P(R) P_R(E) + P(AR) P_{AR}(E)} = \frac{(.70) (.90)}{(.70) (.90) + (.30)(.20)} = \frac{(.63)}{(.63) + (.06)} = 0.91$$

But once the cognitive scientific evidence is received, it might be plausible to raise  $P_{AR}(E)$  to .90. Then the numbers will become:

$$\frac{(.70) (.90)}{(.70) (.90) + (.30) (.90)} = \frac{(.63)}{(.63) + (.27)} = 0.70.$$

Thus, an increase in the conditional probability of essentialist intuitions given anti-realism from .20 to .90 results in a decrease of the posterior probability of realism from 0.91 to 0.70. This is a non-negligible reduction in the probability of realism, due to the cognitive scientific evidence.

What about the contribution of the evidence from physical science? A perspicuous way of handling this is in terms of the prior probability of the truth of kind essentialism; or rather, the posterior probability of essentialism being true given the new evidence about biology and chemistry Leslie introduces. That too would reduce the probability of the truth of the realist hypothesis. Changing the prior probability (not here equated to the "Ur-probability") from .70 to .30, for example, yields the following final posterior probability for realism:

$$P_E(R) = \frac{(.30) (.9)}{(.30) (.90) + (.70) (.90)} = \frac{(.27)}{(.27) + (.63)} = .30$$

Now the posterior probability of realism is reduced all the way to .30. Clearly, a combination of both cognitive and physical scientific evidence can have a significant impact, as we would expect.

#### 6. Fourth Illustration: Theism

The existence of God is a central topic in metaphysics, spanning both the history of philosophy and contemporary philosophy. Traditionally, the positive arguments are predominantly deductive and a priori, the ontological and cosmological arguments being the most prominent examples. It is dubious, however, whether such arguments are terribly promising. Philosophers like them, of course, but it is questionable whether anybody else is blown away by them. It is easy, of course, to produce a deductively valid argument for any specified conclusion. It is harder to produce a sound argument, and even harder yet to produce a sound argument for which the entire targeted audience believes the premises justifiedly. Genuine "proofs" of the existence of God, however, require some such justificational element. Historically, an influential empirical argument for God's existence in which the premises were regarded as justified (for almost anybody) is the argument from design. William Paley's version of this argument was highly persuasive in the early 19th century. However, the evolutionary theory developed by Darwin and his successors has greatly dampened the persuasiveness of this argument. Variation and natural selection seem eminently viable as the real origin of ostensibly designed features of the world.

I shall therefore focus on a distinct type of argument from an observable phenomenon, the extremely high incidence of theistic belief throughout known human history and culture. An ancient and widely invoked criterion of truth (at least in antiquity) was the consensus gentium ("agreement of the people"), which states: "that

which is universal among men carries the weight of truth". Aristotle appealed to this kind of criterion when invoking the opinion of the many. This idea might be re-used in the realm of theism. How could we express this approach in Bayesian terms?

The two hypotheses in the present case are "God exists" (G) and "God does not exist" ( $\sim$ G). Bayes' theorem implies that if we start with roughly equal prior probabilities for these hypotheses, the main driving force in determining a rational posterior credence will be the likelihood ratio (or fraction), that is, the likelihood of there being the evidence in question conditional on G divided by the likelihood of there being this same evidence conditional on  $\sim$ G. The evidence, once again, is the widespread incidence of belief in God (W). The two likelihoods are:

- (1)  $P_G(W)$  and
- (2)  $P_{\sim G}(W)$ .

If  $P_G(W)$  substantially exceeds  $P_{\sim G}(W)$ , then the likelihood quotient exceeds 1.0 and the occurrence or presence of W supports G. If  $P_{\sim G}(W)$  exceeds  $P_G(W)$ , then the occurrence of W supports  $\sim$ G. One might initially be inclined to think that  $P_{\sim G}(W)$  must be extremely small. Why would there be widespread belief in God if God didn't even exist? He would not be there to bestow favors on humankind, to answer prayers, to reveal himself to various prophets, etc. Given the small value of  $P_{\sim G}(W)$  compared with  $P_G(W)$ , the fact that there is widespread belief should generate high credence in God's existence.

This is where possible psychological explanations of belief in God's existence become relevant, in straightforward analogy to the essentialism case. If there are strong psychological propensities toward theistic belief that would operate to produce such belief even in the absence of God's existence, then the value of  $P_{\sim G}(W)$  might not be tiny at all. It might be quite substantial, perhaps even greater than the value of  $P_G(W)$ . As it happens, evidence for such strong propensities toward theistic belief are emerging from various psychological investigations.

Historically, Freud postulated a pathological need for a father figure to explain the prevalence of a God belief and Marx explained it in terms of indoctrination on the part of the powerful. Recently, books by various cognitive scientists, social scientists, and philosophers have pursued other potential explanatory themes (e.g., Atran, 2004; Boyer,

2001; Dennett, 2006). New evidence from cognitive science has also been accumulating. I will now review some of this evidence.

One theme floated by Boyer (2001) is called a "hypertrophy of social cognition": a willingness to attribute psychological states even when it is (evidentially) inappropriate to do so. The classic demonstration of this human tendency was the work of Heider and Simmel (1944). They made a film clip depicting geometrical shapes (circles, squares, triangles) moving about in systematic ways. People watching the film clip instinctively described these shapes as people -- e.g., bullies, victims, heroes -- who have various desires and goals. Young children also interpret such films in this fashion. Bloom & Veres (1999) show that similar results can be obtained even without bounded figures, simply with moving dots and groups of figures such as swarms of squares. In short, interpreting stimuli in terms of goals and purposes seems to be wired into our cognitive engine. That would explain our propensity to postulate unseen creatures like God, devils, sprites, etc. with purposes and designs, even in the absence of probative observational evidence. Similarly, Guthrie (1993) showed that people are inclined to attribute human characteristics to a wide variety of entities with scarcely any trace of humanity, for example, airplanes, automobiles, bags, bells, and even rocks. Babies have similar tendencies -- ostensibly innate -- to attribute intentional characteristics (Bloom, 2007).<sup>12</sup>

More generally, researchers on mindreading, or "theory of mind," have found substantial patterns of error, or "bias," in people's attribution of mental attitudes (mindreading). The best known general type of error is "egocentricity," in which people tend to project (or assign) their own current states to their targets, even when they have no reason to suppose that their targets are similarly situated. The best known bias of this kind goes by the label "curse of knowledge" because it involves people imputing their own knowledge or belief states to others when they have evidence to the contrary (Camerer et al. 1989). Similar tendencies have been demonstrated for egocentric projection of feelings and valuations.<sup>13</sup>

A new wrinkle in this territory, where both excess and misdirection of mentalization are found, led Ara Narenzayan et al. (2012) to conclude that theism is driven by normal people's excessive propensity to assign mental states. Narenzayan began with the well-documented finding that autism involves a deficit in "theory of

mind," or mental state attribution. This deficit predicts that autistic individuals should be less inclined to imagine deities with mental states than normal people are, and hence they should have a diminished inclination to believe in such deities. This hypothesis was tested in a controlled study in which autistic adolescents were found to have less belief in God than did matched neuro-typical controls. Moreover, it was found that levels of mentalizing mediated this relationship. In other words, mentalizing deficits are a "pathway" to disbelief in God. This evidence lends interesting support to the thesis that mentalizing "excesses" (in the population of normals) are heavily responsible for theistic belief, which takes hold in the absence of direct observational evidence.

Admittedly, this psychological evidence is less than overwhelming. It is an interesting and suggestive set of data points. But our purpose here is only to illustrate how cognitive scientific evidence -- if and when it is sufficiently strong -- can have an evidential bearing on the metaphysical issue. However, even if we grant for argument's sake that belief in God is (substantially) the product of run-away mentalizing, how would this affect the Bayesian calculus? As previously explained, it would affect the calculus by boosting  $P_{\sim G}(W)$  to a level considerably higher than one might expect "a priori". This boost non-trivially affects the posterior credence that should be assigned to  $\sim G$ .

One doubt or question may occur to the reader concerning the argument presented to this point. A reader may feel that our argument -- including all four examples -- rests on little more than the observation that everything is potentially relevant to everything else. But haven't we known this since Duhem, and Quine's frequent reminders of the Duhemian thesis?<sup>14</sup> In light of this familiar fact, doesn't my push for the importance of cognitive science have to be buttressed by the demonstration of a special evidential role for cognitive science? Otherwise, why think that cognitive science deserves anything but an incidental and marginal role in the methodology of metaphysics? What is special about cognitive science that we should single it out as particularly important to metaphysics?

A partial answer -- well-illustrated by the four examples above -- is that metaphysicians often appeal to mental states as "data" for purposes of metaphysical theorizing. These mental states include beliefs, intuitions, experiences, and behavioral propensities. Within the class of behavioral propensities are verbal propensities to

classify certain scenarios one way or another, e.g., as instances of causation or non-instances of causation. Cognitive science has much to say about all of these types of states and behaviors, including theories to explain and systematize them. Since such explanatory theories can either cohere or conflict with metaphysical theories (e.g., theories of causation), cognitive scientific perspectives have a distinctive bearing on metaphysical theory. This should have been especially clear in the examples of temporal flow and of theistic belief.

## 7.. Cognitive-Systems Relationalism: The Case of Color.

Thus far my illustrations of how cognitive science can contribute to metaphysics tend to favor anti-realist over realist metaphysical theories. How, one might wonder, is this compatible with my earlier espousal of a realist (as contrasted with conceptualist) approach to metaphysics? The problem here arises from the confusing terminology that has grown up in metaphysics (or metametaphysics). "Realism" is a notoriously elusive term to define, and different people use it in different ways. Without aspiring to completeness, let me say a few things about how I understand these terms and their relations to one another. The examples to be discussed in these final two sections will then fill out my picture via the theories I wind up endorsing (either fully or tentatively).

The sense of "realism" in which I have a high commitment to it is the sense in which it is roughly equivalent to "objectivism." Metaphysics should seek what is really or objectively true about the nature of things as opposed to what people may naively, commonsensically, or uncritically think about the nature of things. This does not, of course, imply that realism entails mind-independence in the sense that metaphysics should never be willing to characterize the nature of anything (including the nature of any property) in mentalistic or psychologistic terms. On the contrary, certain properties may be best characterized in precisely such ways. The property of color, for example, may be best spelled out in terms of relationships with kinds of mental states or kinds of cognitive systems. How does the taxonomical terminology of "realism" versus "anti-realism" relate to this point? Unfortunately, it is completely orthogonal to it. Anti-realism connotes a type of philosophical theory that is revisionary, at variance (in one or more ways) with

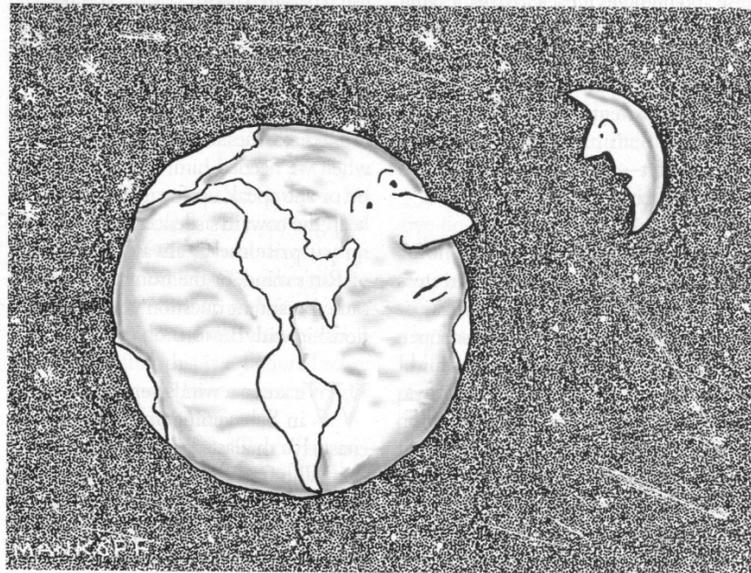
naïve or commonsensical ways of thinking about the target phenomena. But this in no way implies a resistance to all matters mental, psychological, or cognitivist. Also, anti-realist approaches do not reject all uses of the mental in the execution of metaphysics. Although anti-realists may be prepared to distance themselves from what is commonly thought, they need not deprive themselves of reliance on intuitions and experiences as prima facie evidence concerning the phenomena under investigation. How else can the theorist get started in picking out our quarry when we select color as the property (or relation) we mean to investigate? We must rely on our initial grasp of color phenomena, which includes the undergoing of color-perceptual experiences. This is compatible with the possibility that the fundamental nature of color, understood in a suitably "objective" and scientific fashion, is a complex phenomenon by no means exhausted by our initial familiarity with it. A child can recognize lightning by a characteristic visual experience.. But the real (underlying) nature of lightning will radically outstrip what the child understands about it. Since the notion of realist metaphysics (as understood here) presupposes a convergence with scientific metaphysics, it should not be surprising that particular instances of metaphysics would invite cognitive scientific notions to play important roles. The final two sections of this paper illustrate this idea, using color and persistence over time as the metaphysical properties (or relations) of interest.

In this section I work my way toward a relationalist approach to the metaphysics of color. Relationalism -- more specifically, cognitive-systems relationalism (C-relationalism) -- is the second "template" I wish to advance in this paper as another type of example of what cognitive science can contribute to metaphysics. CS-relationalism claims that for certain properties or phenomena of metaphysical interest, the optimal ontological construal of their nature involves an added relatum that is not naively apprehended, the need for which is cognitive-scientifically well motivated. Moreover, the objects that are the potential values of this extra parameter (or "addicity") are themselves cognitive systems.

Relationalism about color was first formulated by E. W. Averill (1992) and considerably expanded more recently by Jonathan Cohen (2004, 2009). Cohen regards this view as a variety of realism (certainly as an alternative to irrealism). Let me expand a bit about what I mean by objectivism (which I equate with one construal of realism). It

seeks to identify the objective features of the universe, where "objective" (in my terminology) connotes the truth about things as viewed from a universal or cosmic point of view, not merely a local or partial perspective. Realist or objectivist theories try to avoid anything that smacks of anthropocentrism, species chauvinism, or the like. This aim is shared by Cohen's relationalism about color, as we shall see.

A naïve theory of color is one that affirms and/or explicates our naïve understanding of color, and regards this as an adequate metaphysical understanding of the property itself. A good example of this is the conception of color Mark Johnston (1992) calls "revelation". This conception holds that the intrinsic nature of a color is fully revealed by a standard visual experience of it. Since the kind of visual experience tacitly referred to here is human experience, this implies that human color perception is fully revelatory of color's true nature. To the extent that relationalism aspires to a more objective, universal perspective on color, it seeks to avoid such species chauvinism. It especially seeks to avoid unwitting chauvinism, a recurrent danger in customary modes of doing metaphysics to the extent that it relies heavily on human intuitions and experiences. (See figure 3 below, from the New Yorker).



*"What's so galling is that you don't even realize how Earthist you are."*

Cohen (2004: 453-454) contrasts color relationalism and nonrelationalism as follows. Nonrelationalists typically understand color as a mind-independent or perceiver-independent phenomenon. A typical form of nonrelationalism holds that colors are a kind of physical property, e.g., a reflectance property of surfaces, a transmittance property of transparent surfaces and volumes, or an emittance property of luminous sources. By contrast, relationalism typically holds that colors are constituted in terms of relations between objects and perceivers (as well as values of other parameters such as viewing conditions). Color dispositionalism, for example, holds that colors are dispositions to cause certain sensations in certain kinds of minds.

Which kinds of minds? Red is commonly characterized as a disposition of object surfaces to reflect light in ways that cause us -- i.e., normal human perceivers -- to have a certain kind of visual experience. But can all color phenomena be specified in terms of relations between object surfaces, light, and human experience? What about distinctive color phenomena associated with what other species experience, especially species with different types of visual systems? As comparative color science shows, pigeon visual systems are tetrachromatic: they use four (functionally individuated) channels to encode color information. Therefore, an arbitrary color can be perceptually matched for a pigeon by a linear combination of four appropriately chosen primaries. In contrast, normal human visual systems are trichromatic: they use three (functionally individuated) channels to encode color information. An arbitrary color can be perceptually matched for a human being by a linear combination of three appropriately chosen primaries. As Cohen explains (2004: 462), because of this difference there are pairs of surfaces that are perceptual matches for human visual systems but are not matches for pigeon visual systems. The mantis shrimp has at least ten or twelve kinds of cones in its visual receptive system (Cronin and Marshall, 1989), so its color responsiveness is presumably many times more complicated and variegated than the human system.

Cohen proceeds to pose a dilemma for color science and metaphysics. The central question is: which of several alternative perceptual variants veridically represents the color of a given surface? He weighs the options as follows:

The four choices in logical space are these: (i) confine our attention to human visual systems, declaring that how things look to pigeon (and other) visual systems is irrelevant to the colors of objects; (ii) defer to the pigeons, holding that the way things look to them determines the true colors of objects; (iii) declare that neither we nor the pigeons are the true arbiters of color, and instead select a different standard; (iv) adopt the ecumenical policy that both sorts of visual systems are right, and that one and the same object can have more than one color property. (Cohen, 2004: 462)?

Alternative (iv) is relationalism, which can be explained as the relativizing of colors to types of visual systems. Of course, this is not the only bit of relativization needed for an adequate theory of color. Along with many other color philosophers and scientists, Cohen argues that the theorist needs multiple parameters, each of which can take many values, to provide enough relativization. However, relativization to types of visual systems is the most relevant kind for present purposes, so I highlight it in my presentation.

As Cohen expresses it, colors are constituted "in terms of relations to ... kinds of visual systems... It is a consequence of this relational construal that one and the same object can be simultaneously green for your visual system and not green for the visual system of the pigeon on your window ledge" (2004: 463). I now add the obvious point that color science, including the cognitive part of color science, is what we need in order to supply the right values of visual-system parameters in order to obtain correctly relativized facts. Once suitable relativizations are in place, the facts in question would not be anthropocentric facts, or species-chauvinist facts. They would be cosmic facts, which hold from the "perspective" of the universe. This is the kind of fact sought by objectivism, or realism. Supplying such a fact -- in detail, at any rate -- hinges partly on cognitive science, because visual systems are proper parts of cognitive engines (which cognitive science studies).

Although the relationalist theory of color is (or aims to be) realist in the sense indicated above, it is also clearly anti-realist in the sense associated with the realism/anti-realism terminology. Ordinary people untutored in color science do not understand or

represent the color red as a complex relational affair. It is more likely understood as a simple, monadic property of object surfaces. But color relationalism does not purport to be a theory about how red and other colors are commonsensically understood, or (mentally) represented. It is a theory about the nature of color (or colors). Thus, there is no inconsistency between calling relationalism a species of realism and yet also a kind of anti-realist theory.<sup>15</sup>

According to CS-relationalism there is no unique answer to the question of what color a given object or surface is; one must first relativize the question to a specified type of visual system (as well as viewing conditions, contrast conditions, etc). Then an answer is forthcoming (in principle).<sup>16</sup> Of course, ordinary thought and talk make no explicit mention of the parameters that relationalism invokes to constitute the colors. But such parameters, Cohen argues, are tacitly filled in by generalizing from our own case or the cases of organisms like us (2004: 471). When we say that x is green simpliciter we mean that x looks green to visual systems like our own and in viewing conditions like those we typically encounter. This is not an adequate general account of what color is, however, from a neutral and universal point of view. CS-relationalism provides a genuinely objective characterization of colors that holds independently of any particular species or visual system. The relational truths in question are truths "for" any cognizer capable of grasping them even if those cognizers do not themselves instantiate the relations in question. (They might be members of a species with no color-vision system at all.). In this sense, the theory does not "privilege" the position of one kind of visual system over others. This kind of story, of course, is attractive to science, the mission of which is to seek truths that can be appreciated by earthlings and aliens alike. Such truths are species-independent rather than anthropocentric or otherwise species chauvinistic.

It is also a story to which science -- including cognitive science -- has contributed. This is no news to philosophers of color. They have been attuned to the deliverances of color science at least since the pathbreaking work of C. L. Hardin (1988).<sup>17</sup> But even though the relevance of cognitive science to color metaphysics is well known to philosophers of color, no analogous situation holds for other branches of metaphysics. The message of the present paper is therefore still timely. Section 8 moves to a different metaphysical terrain and argues (tentatively) for CS-relationalism in that branch of

metaphysics, which has thus far featured rather little penetration by cognitive science (at least within the mainstream metaphysical literature<sup>18</sup>).

#### 8. Cognitive-Systems Relationalism: The Case of Persistence

Assume that a metaphysician wishes to state the truth-conditions for the cross-temporal identity (or unity) of a pair of objects (or object-stages). The aim is to complete a schema of the following form:

Objects x and y at spatiotemporal locations L and L' respectively are numerically identical with one another (or bear the unity-relation to one another) if and only if ----."19

Further assume that the metaphysician embraces the ubiquitous methodology of at least beginning from evidence of people's intuitive judgments about particular cases of identity through time. A theory that fails to conform substantially to this intuitional or judgmental evidence, or fares worse than its rivals, would face major obstacles. It would have a lot to "answer for" in the eyes of metaphysicians.

Does cognitive science have a contribution to make here? One possible contribution (rarely exploited) is to unearth evidence of persistence judgments that are not readily accessible to the subject himself or to casual observers. These judgments might be neither verbal nor explicit. But they might be detectable by inference from psychological experimentation. Such judgments might be extremely important components of a subject's wider thought and cognition. They might even lie at the foundation of other exercises of people's ability to track objects through time. Finally, establishing the existence and contents of such judgments, as well as how they are executed, might pose challenges to prevailing metaphysical theories about the persistence relation. Here is one such example.

A number of philosophers have held that sortals, or kind terms, play a pivotal role in judgments of identity over time. Indeed, some have held that identity over time is always relative identity, in the sense that such judgments must always be relativized to a

sortal (Geach 1967; Griffin 1977). They must have the form "b is the same F as c" for some suitably chosen sortal F.

We can now report evidence from cognitive science that seems to cut against this sortal-based view of persistence, at least as a necessary component of persistence judgments. In particular, this evidence supports the notion that at a certain level of cognition an object's identity over time is computed without appeal (or only minimal appeal) to sortals. If this description is accurate, it is a case of cognitive-scientific findings having an evidential bearing on a metaphysical thesis about persistence. The evidence to which I allude does not merely show that there are occasional or unusual judgments of the kind of question, but that they are made continually during waking life without awareness that one is doing so.

The beliefs or judgments in question are a feature of a sub-system of the visual system.<sup>20</sup> This sub-system is called the "mid-level object-file system," The term "mid-level" refers here to an intermediate level of cognition between low-level sensory processing and fully conceptual processing. An "object file" -- a phrase introduced by Kahneman and Treisman (1984; cf. Kahneman et al., 1992) -- is a visual representation that sticks to a moving object over time, storing and updating information about that object's surface features. In other words, it tracks or continually re-identifies an object even while the object's properties (e.g., color, size, shape, and kind) change. Here is how Flombaum, Scholl, and Santos (2009) explain the problem facing the visual system and how it appears to address the problem.

The impoverished and rapidly changing stimulation on the retina looks very different from the stable world of discrete persisting objects that populate our visual experience. To get from the features on the retina to the objects that we experience, the visual system must solve several correspondence problems. One of these problems has to do with sameness: the visual system must decide whether each bit of stimulation reflects an object that has already been encountered (which might occasion the updating of an existing object representation) or a new one (which might occasion the creation of a new object representation).... The study of object persistence [has] converged on a core principle that guides the creation

and maintenance of persisting object representations: the principle of spatiotemporal priority. When identifying objects as the same individuals over time, the visual system appears to rely on their spatiotemporal histories -- that is, when, where and how they were encountered -- to a greater extent than their visual surface features. (Flombaum et al., 2009: 135)

In this literature, researchers talk about "file folders" being created or destroyed and various information being added to the file folder, where a given file folder represents a single object persisting over time.

What is crucial about this "mid-level object-file" theory is that the system works (mainly) by attending to an object's spatiotemporal properties, but not its sortal properties. The system tracks an object over time not by considering its shape, its color, or the kind of thing it appears to be, but what positions it occupies as it changes locations over time. Here I depart slightly from what is actually found in the research literature and consider what might be found (though it is very close to the actual findings). This is warranted by our interest in possible findings of cognitive science, not merely actual findings. (We are interested in what cognitive science is capable of contributing to metaphysics, not simply what we know at the present time it is prepared to contribute.) So suppose it were clearly stated (and this might indeed be the case, as suggested by later experiments to be reported) that the visual system continually makes persistence judgments that do not involve sortals, contrary to what the sortal-based theories of Geach and Griffin require.<sup>21</sup> This would be a clear refutation of that a sortal-based philosophical theory, understood as an unqualified generalization about the mind's method of tracking identity over time. It would be especially significant because the mid-level object-file system is fully in place in infancy, by at least ten months of age (Xu and Carey, 1996; Spelke, Kestenbaum, Simons, and Wein, 1995). Here, then, would be cognitive scientific evidence with direct (and conclusive) bearing on a particular metaphysical thesis.

Complementary evidence for the properties just mentioned of the mid-level object-file system comes from a separately developed theory called "visual indexing," due to Zenon Pylyshyn (1989, 2004; Pylyshyn and Storm 1998). These investigators

postulate a mechanism whereby object-based individuation, tracking, and access are realized. In an experimental paradigm for studying visual indexing called "multiple object tracking," or MOT, participants are shown a number of qualitatively identical objects on a screen. A few of these objects are flashed to indicate that they are the targets to which participants should attend. In the next phase of the session, all objects on the screen are set into random, unpredictable motion about the display, often changing their color, size, shape, and kind. When they stop moving, participants are asked to re-identify the objects originally shown as targets. This sounds like a difficult task, but people are surprisingly adept at re-identifying members of the highlighted set (the targets) despite their continually crisscrossing one another and changing color, size, shape, and kind during motion. All this suggests that the system attends only to location, movement, and other such spatiotemporal properties while ignoring sortal properties. This is how the researchers interpret what transpires, and it is consistent with other evidence about the computational features of the mid-level object-file system.

Thus far, we have illustrated how cognitive scientific evidence proves relevant to past debates in the metaphysics of persistence. But this demonstration of evidential relevance is only of a piece with the first half of this paper. It does not address the theme presented in section 7, concerning CS-relationalism. What do we say on this subject?

The next thing to be reported is a conclusion by relevant cognitive scientists that there is a second system for persistence judgments, independent of the mid-level object-file system, which does use sortals or kinds.<sup>22</sup> Apparently, this is the system people draw upon when deciding that a person ceases to exist when she dies, or that a car ceases to exist when it is crushed into a cube of metal at the junkyard, despite the fact that there is plenty of spatiotemporal continuity between the person's body and corpse in the first case and between the car and the metal cube in the second case.

Differences between the object-file system and the kind-based system can be illustrated with the help of Figure 3 below (from Carey and Xu, 2001). Each of the two systems can be launched into operation in connection with this figure. Suppose first that you examine Panel 1 but then lose perceptual contact with it and return five minutes later to view Panel 2. How would you describe what has happened? You would probably say that the rabbit has moved from above and to the left of the chair to below and to the right

of it, while the bird has moved from the bottom left to the top right. In other words, you would re-identify the rabbit in Panel 1 with the rabbit in Panel 2, despite their difference in location. Numerical identity would be dictated by kind membership.

Next imagine that a fixation point replaces the chair and Panels 1 and 2 are projected rapidly one after the other onto a screen you are viewing, while you maintain fixation on the common fixation point. (This is a kind of visual display that is well-studied in the perception literature, so vision scientists know exactly how it will be seen by normal perceivers.) If the timing of the rapid change process supports what is -called "apparent motion," what will your perception be? You will see two individuals each changing back and forth between a white bird-shaped object and a black rabbit-shaped object as they move side to side. The visual system so computes the numerical identity of the objects' undergoing apparent motion as to minimize the total quantity of movement. This system rarely takes sortal or kind information into account -- only when spatiotemporal considerations are equated. Using evidence like this, Carey and Xu (2001) conclude that the kind-based object individuation system is architecturally distinct from the mid-level object-file system.

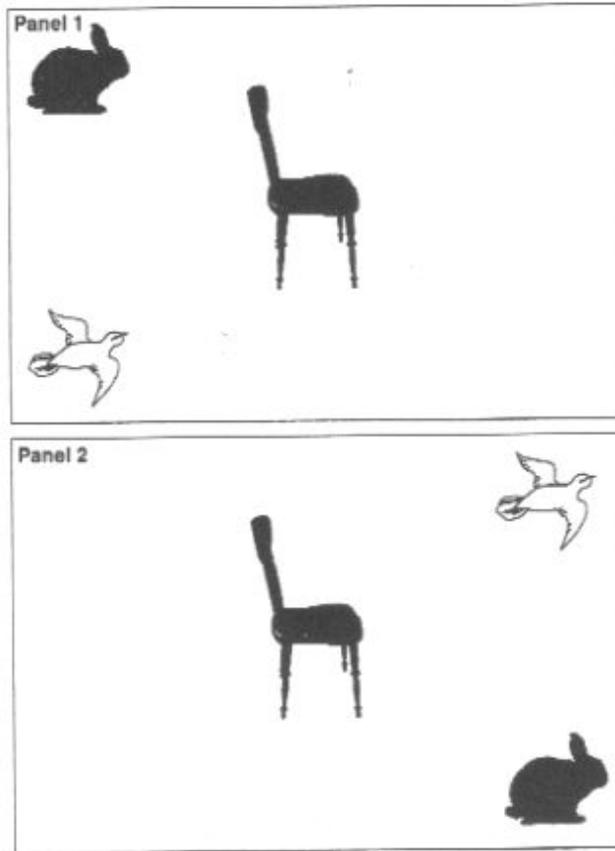


Figure 3

Suppose now that a metaphysician's aim is to formulate principles constitutive of object persistence, principles articulating standards for the correctness, or veridicality, of such judgments. The looming problem is how to get a consistent and unified set of principles that reflect the operational features of the two psychological systems. They seem to be in flat-out contradiction with one another. How can a single self-consistent set of principles be formulated for the metaphysics of persistence? Self-consistency is required for a satisfactory set of metaphysical principles, but how can this be attained if one has to honor the operations of both systems, which are quite different from one another? Adopting criteria that reflect the operations of both systems would seem, inevitably, to impose conflicting criteria. There seems to be no (consistent) way to merge the operational principles of both systems. So how could objective persistence facts be determined by two such conflicting sets of principles?

This raises the question whether, in light of the existence of two different identity-judgment systems, we can continue to assume that there are any objective facts of persistence at all, or that such facts can reflect (all of our) naïve judgments of persistence.. If the two systems' basic operating principles are so different, can the metaphysician continue to rely on the reliability and evidential probity of their judgmental outputs? If not, which set of judgments should be deemed more reliable and trustworthy? How should we select the more trustworthy indicators of the metaphysical facts? A metaphysical and/or methodological conundrum seems to emerge here.

There are several ways that a metaphysician studying might respond. The first response I shall call the two-families-of-propositions response. According to this response, each psychological system endorses a distinct family of propositions. The canonical form of one of these families involves propositions that assert (or deny) numerical identity (or unity) relations between two objects with no third term for sortals. The canonical form of the second family involves propositions that assert or deny a ternary relation between pairs of objects plus a sortal. Call these propositions "two-place" and "three-place" propositions respectively. Evidently, even when a proposition in one family makes an assertion about the same pair of objects as the other does, they will convey different meanings, because the two relations in question must differ in meaning. In effect, the metaphysician is trying to diagnosis the original problem by saying that

there is a "verbal dispute" between the sortalist and non-sortalist (cf. Hirsch, 2005). Sortal theorists and non-sortal theorists of persistence are simply using words differently. This approach does have the capacity to eliminate the original problem of inconsistency. But whether one is warranted in tracing the original problem to a verbal dispute is questionable. Certainly the mid-level object-file system is not a verbal system at all.

A second type of response is to insist that only one family of judgments is an "authoritative" family. Only judgments issued by the designated system are genuinely reliable indicators of the metaphysical facts of persistence. Call this the privileged cognitive system response.<sup>23</sup> What rationale might be offered for it, and what criteria for selecting the authoritative system should be used? The correct system, it might be argued, is the sortal-based system (supplemented, as it surely must be, by considerations of spatiotemporal continuity). The reason is that the two systems currently under discussion exemplify a well-known pattern described by so-called dual-process theorists (cf. Evans and Frankish, 2009). The mid-level object-file system is a paradigm case of a system-1 process in being evolutionarily old, unconscious, automatic, fast, etc., whereas the kind-based system is a paradigm case of a system-2 process in being evolutionarily recent, conscious, controlled, slow, etc. Reflecting on other dual-process pairs, it is noteworthy that system-1 members of the pair are typically more biased and inaccurate than their system-2 counterparts. System-1 processes may be viewed as mere heuristics for system-2 processes, and heuristics are known to be biased and error-prone.

A third type of response available to a persistence metaphysician is a version of CS-relationalism that mirrors CS-relationalism for color. Under this approach, no priority is assigned to either cognitive system as the determinant of truth conditions for persistence. Rather, we should craft truth conditions for re-identification that include relativity to cognitive systems, much as CS-relationalism for color includes a parameter for color-vision systems. Although human beings have two systems for persistence judgments -- according to the cognitive science literature summarized earlier -- other creatures might have still different types of persistence systems. On what basis are human systems (or either one of the human systems) to be "privileged" over the systems of other creatures? An avoidance of arbitrariness is part of what drives Cohen toward relationalism about color, and it is one of my preoccupations here as well.

As explained in the previous section, I consider it an aim of metaphysics to characterize the world in "cosmic" terms. No prioritizing of "local" perspectives is allowed. Keeping the cosmic perspective in mind, can we really credit the notion that the human object-file system or the human kind-based system regularly generates the uniquely "correct" or "right" persistence judgments whereas other systems with even modest departures from these would go "wrong"? Would any different system inevitably fail to carve persistence at its joints? This is analogous to saying that only the human visual system represents color facts correctly, whereas pigeons, dogs, and mantis shrimp are fated to get the color facts systematically wrong. Such anthropocentrism is difficult to credit.

CS-relationalism can come to the rescue. It starts from the assumption that there are no unrelativized facts of the matter pertaining to object persistence. But it then finds facts of a sort by relativizing the truth conditions of persistence to this or that selected cognitive system, which could usefully be the user's own system (or one of these systems, in case there are two or more). By introducing this added parameter, CS-relationalism makes room for a new genre of persistence facts, which are perfectly objective in the sense we have emphasized: universal, cosmic, non-species chauvinistic. It would take the following form.

(CSPERS) Objects  $x$  and  $y$  (at spatiotemporal locations  $L$  and  $L'$  respectively) are object-unified) relative to cognitive system CS1 if and only if ---.

I shall not make a categorical choice between these alternative responses (although I do lean toward CS-relationalism). However, I now submit that here we find another prototype of a way in which cognitive science can contribute to metaphysics. By uncovering cognitive processes that are unavailable to introspection we get a much more nuanced understanding of the complex cognitive activities that go on "below the surface". These activities can pose puzzles or conundrums that confront the metaphysician with previously unimagined choices. The scientific revelations will not themselves be capable of resolving the newly posed metaphysical puzzles, but the metaphysics is bound to improve to the extent that it recognizes and grapples with the more complex array of

choices. A scientifically improved understanding of color experience and persistence judgments presents a different kind of window on the generation and resolution of metaphysical problems that naturally go hand-in-hand with advances in cognitive science.<sup>24</sup>

## References

- Atran, S. (2004). In Gods We Trust. The Evolutionary Landscape of Religion. New York: Oxford University Press.
- Averill, E.W. (1992). The relational nature of color. Philosophical Review 101: 551-588.
- Bloom, P. (2007). Religion is natural. Developmental Science 10(1): 147-151.
- Bloom, P. and Veres, C. (1999). The perceived intentionality of groups. Cognition 71: B1-B9.
- Boyer, P. (2001). Religion Explained. New York: Basic Books.
- Byrne, A. and Hilbert, D.R. (eds.) (1997a). Readings on Color. Volume 1: The Philosophy of Color. Cambridge, MA: MIT Press.
- Byrne, A. and Hilbert, D.R. (eds.) (1997b). Readings on Color, Volume 2: The Science of Color. Cambridge, MA: MIT Press.
- Cohen, J. (2004). Color properties and color ascriptions: A relationalist manifesto. Philosophical Review 113(4): 451-506.
- Cohen, J. (2009). The Red and the Real. Oxford: Oxford University Press.
- Carey, S. (2009). The Origin of Concepts. Oxford: Oxford University Press.
- Carey, S. and Xu, F. (2001). Beyond object files and object tracking: Infant representations of objects. Cognition 80: 179-213.
- Chomsky, N. (1980). Rules and Representations. New York: Columbia University Press.
- Craig, A.D. (2002). How do you feel? Interoception: The sense of the physiological condition of the body. Nature Reviews Neuroscience 3: 655-666.
- Craig, A.D. (2010). Interoception and emotion: A neuroanatomical perspective. In M. Lewis, J. M. Haviland-Jones, and L. Feldman Barrett (eds.). Handbook of Emotions, 3rd edition (pp. 272-288). New York: Guilford Press.
- Cronin, T.W. and Marshall, N.J. (1989). A retina with at least ten spectral types of photoreceptors in a mantis shrimp. Nature 339(6220): 137-140.
- Dennett, D.C. (2006). Breaking the Spell: Religion as a Natural Phenomenon. New York: Viking.

- Evans, J.S.B.T. and Frankish, K. (eds.) In Two Minds: Dual Processes and Beyond.  
Oxford: Oxford University Press.
- Flombaum, J.I., Scholl, B.J., and Santos, L.R. (2009). Spatiotemporal priority as a  
fundamental principle of object persistence. In Hood, B.M and Santos, L.R. (eds.)  
The Origins of Object Knowledge (pp. 135-164). Oxford: Oxford University  
Press.
- Geach, P.T. (1967). Identity. Review of Metaphysics 21: 3-12.
- Gelman, S.A. (2003). The essential child: Origins of essentialism in everyday thought.  
New York: Oxford University Press.
- Gregory, R.L. (1966). Eye and Brain. London: Weidenfeld & Nicolson.
- Goldman, A.I. (1987). Cognitive science and metaphysics. Journal of Philosophy 84:  
537-544.
- Goldman, A.I. (1989). Metaphysics, mind, and mental science. Philosophical Topics 17:  
131-145.
- Goldman, A.I. (1992). Cognition and modal metaphysics. In Goldman, Liaisons:  
Philosophy Meets the Cognitive and Social Sciences (pp. 49-66). Cambridge,  
MA: MIT Press.
- Goldman, A.I. (2007). A program for "naturalizing" metaphysics, with application to the  
ontology of events. The Monist 90(3): 457-479.
- Goldman, A.I. (2012). A moderate approach to embodied cognitive science. Review of  
Philosophy and Psychology 3(1): 71-88.
- Gregory, R.L. (1966). Eye and Brain. London: Weidenfeld and Nicolson.
- Griffin, N. (1977). Relative Identity. Oxford: Oxford University Press.
- Guthrie, S.E. (1993). Faces in the Clouds: A New Theory of Religion. New York:  
Oxford University Press.
- Hardin, C. L. (1988). Color for Philosophers: Unweaving the Rainbow. Indianapolis:  
Hackett.
- Haslanger, S. (2003). Persistence through time. In M. Loux and D.W.Zimmerman  
(eds.), The Oxford Handbook of Metaphysics (pp. 315-354). Oxford: Oxford  
University Press.

- Heider, F. and Simmel, M. (1944). An experimental study of apparent behavior. American Journal of Psychology 57: 243-259.
- Hirsch, E. (2005). Physical-object ontology, verbal disputes, and common sense. Philosophy and Phenomenological Research 70(1): 67-97.
- Johnston, M. (1992). How to speak of the colors. Philosophical Studies 68: 221-263.
- Joyce, R. (2006). The Evolution of Morality. Cambridge, MA: MIT Press.
- Kahneman, D. and Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman and D.R. Davies (eds.), Varieties of Attention (pp. 29-61). New York: Academic Press.
- Kahneman, D., Treisman, A. and Gibbs, B. (1992). The reviewing of object files: Object-specific integration of information. Cognitive Psychology 24: 175-219.
- Keil, F.C. (1989): Concepts, kinds, and cognitive development. Cambridge, MA: MIT Press.
- Kripke, S. (1980). Naming and Necessity. Cambridge, MA: Harvard University Press.
- Ladyman, J. and Ross, D. (2007). Every Thing Must Go: Metaphysics Naturalized. Oxford: Oxford University Press.
- Le Poidevin, R. (2007). The Images of Time. Oxford: Oxford University Press.
- Leslie, S.-J. (2013). Essence and natural kinds. In T. Gendler and J. Hawthorne (eds.), Oxford Studies in Epistemology, vol. 4. Oxford: Oxford University Press.
- Lewis, D. K. (1986). On the Plurality of Worlds. Oxford: Basil Blackwell.
- Maudlin, T. (2007). The Metaphysics within Physics. Oxford University Press.
- McKay, D. (1958). Perceptual stability for a stroboscopically lit visual field containing self-luminous objects. Nature 181: 507-508,
- McTaggart, J.M.E. (1908). The unreality of time. Mind 18: 457-484.
- Narenzayan, A., Gervais, W.M., and Trzesniewski, K.H. (2012). Mentalizing deficits constrain belief in a personal god. PLoS One 7(5): 1-8.
- Paul, L.A. (2010). Temporal experience. Journal of Philosophy 107(7): 333-359.
- Paul, L.A. (manuscript). "Experience and the Arrow."
- Putnam, H. (1975). The meaning of meaning. In H. Putnam, Mind, language and reality. Cambridge: Cambridge University Press, 215-271.

- Pylyshyn, Z.W. (2004). Some puzzling findings in multiple object tracking MOT: I. Tracking without keeping track of object identities. Visual Cognition 11(7): 801-822.
- Pylyshyn, Z. W. and Storm, R.W. (1998). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. Spatial Vision 3: 179-197.
- Rips, L. J., Blok, S, and Newman, G. (2006). Tracing the identity of objects. Psychological Review 113(1): 1-30.
- Schaffer, J. (2010a). Monism: The priority of the whole. Philosophical Review 119(1): 31-76.
- Schaffer, J. (2010b). The internal relatedness of all things. Mind 119: 341-376.
- Scholl, B. J., ed. (2002). Objects and Attention. MIT Press.
- Scholl, B.J. (2007). Object persistence in philosophy and psychology. Mind and Language 22(5): 563-591.
- Schuster, M.M. (1986). Is the flow of time subjective? Review of Metaphysics 39L 695-714.
- Sider, T. (2011). Writing the Book of the World. New York: Oxford University Press.
- Spelke, E.S. (1990). Principles of object perception. Cognition Science 14: 29-56.
- Spelke, E.S., Kestenbaum, R., Simons, D.J., and Wein, D. (1995). Spatiotemporal continuity, smoothness of motion, and object identity in infancy. British Journal of Developmental Psychology 13: 113-142.
- Strawson, P.F. (1959). Individuals: An Essay in Descriptive Metaphysics. London: Methuen.
- Street, S. (2006). A Darwinian dilemma for realist theories of value. Philosophical Studies 27(1): 109-166.
- Vignemont, F. (de) (2011). Bodily awareness. Stanford Encyclopedia of Philosophy.
- White, R. (2010). You just believe that because.... In J. Hawthorne and J. Turner (eds.), Philosophical Perspectives 24: 573-615.
- Xu, F. (1997). From Lot's wife to a pillar of salt: evidence that physical object is a sortal concept. Mind and Language 12: 365-392.
- Xu, F. and Carey, S. (1996). Infants' metaphysics: the case of numerical identity. Cognitive Psychology 30: 111-153.

---

<sup>1</sup> This would not necessarily imply that every metaphysician is personally obliged to consult the cognitive-scientific literature. It would only imply that the community of metaphysicians should be collectively informed of relevant findings and theoretical orientations in cognitive science. The most efficient way to achieve this end might be a division of labor, in which certain metaphysicians are in the vanguard of tracking relevant cognitive-scientific findings and communicating them to other practitioners.

<sup>2</sup> Over a 25-year period I have periodically tried to make the sort of argument undertaken here, in both graduate seminars and publications (Goldman 1987, 1989, 1992, 2007). In the same period a number of other people have advanced the importance of inputs from cognitive science in pursuing various local parts of metaphysics, including color, time, and moral value especially. Nonetheless, casual conversation with prominent contemporary metaphysicians reveals an abundance of skepticism toward the general methodological thesis of the kind advanced here. So efforts at articulation of the general thesis and defense of it can hardly be classified as mere preaching to the choir.

<sup>3</sup> In the past I have sometimes described this as prescriptive as opposed to descriptive metaphysics (Goldman 1989, 2007). The original distinction, of course, is due to Strawson (1959).

<sup>4</sup> There is a problem of detail here, because what "R" represents -- reality as a whole -- already includes the human brain, so there is an unhappy bit of overlap here. (Thanks to Yoaav Isaacs for raising this issue.) For present purposes, however, this messy detail will be ignored, lest it distract us from the main point.

<sup>5</sup> Relevant specifications of realism vary from topic to topic and from author to author. So I won't try to give a general characterization of realism that works for all cases. There is enough unity, however, that the relevant points can be made without a precise definitional commitment.

<sup>6</sup> Metaphysics is not the main target of White's discussion; it is epistemology. But you might say that the primary focus of the present paper is also epistemology, i.e., the epistemology (methodology) of metaphysics.

<sup>7</sup> Orthodox Bayesians do not believe in the updating of likelihood priors. But this is a very natural way to treat the kinds of examples I want to discuss here. Readers who have a strong preference for orthodox Bayesianism can re-do the analysis by updating posterior probabilities of H and ~H with the new evidence.

<sup>8</sup> In more recent work Paul (forthcoming) explores a similar theme with respect to time's "arrow". This additional topic, however, is beyond the present purview.

<sup>9</sup> For an accessible treatment of Craig's research on the neuroanatomy of interoception -- a sense that monitors the physiological condition of the body -- see Craig (2010), though that article focuses specifically on the interoception of emotion.

<sup>10</sup> Admittedly, it is not clear how cognitive science could recognize a moral sense if it encountered one. Even if it could determine circumstances in which the brain represents an action, say, as right or wrong, how would it determine that such a representation is veridical or not? If it cannot make such a determination, how can it tell that a genuine moral sense is at work? However, by the same token, what evidence does the ordinary person or the armchair metaphysician have for thinking that we have anything like a moral sense, or any other method for recognizing good or bad reasons for favoring something?

<sup>11</sup> I encountered Leslie's paper after most of the present paper was complete, including the three other illustrations analyzed in a Bayesian framework. As it happens, I had criticized Kripke's thesis of the necessity of origin in Goldman 1992. That treatment was extremely brief and under-developed, however.

<sup>12</sup> How exactly is the statement to be defended that people attribute mental states of the indicated sorts in the absence of probative evidence? What kind of evidence would be probative? This can be answered in

---

different ways. One answer would draw on a popular account of the nature of mental states, viz., the functional-role account. If being in a desire state or a goal state with a certain content is a matter of having dispositions to interact with environment inputs, behavior outputs, and also lawlike relations between a goal state and other kinds of internal states, then it appears that none of the attributions cited in the foregoing studies were based (even remotely) on solid evidence. They just seem (in many cases) to be based on lively movements and little else.

<sup>13</sup> See Goldman (2006: 164-73) for a review of the psychological literature on egocentric biases.

<sup>14</sup> Thanks to Jonathan Schaffer for expressing this worry.

<sup>15</sup> See Cohen (2009: 138-147) for broadly similar views, although we may not agree in every detail.

<sup>16</sup> In light of these facts, it is odd to find Sider (2011: 1-2) choosing a color example to support his insistence that proper metaphysical realism should carve nature at its joints. I don't necessarily disagree with Sider about the joint-carving goal of metaphysics, only with his assumption that unrelativized color classifications are a particularly clear example of successful joint-carving.

<sup>17</sup> See Byrne and Hilbert (1997a, b).

<sup>18</sup> I make this qualification because there has been a substantial convergence of interests among both psychologists and philosophers on certain aspects of the metaphysics of persistence. Most of the direction of flow, however, has been from philosophy to psychology. Few philosopher-metaphysicians have made important use of psychological findings, especially in a metametaphysical vein. On the other hand, one can find a great deal of discussion of philosophers' debates about persistence (and other topics on physical-object ontology) by psychologists. Two particularly extensive and detailed treatments are Rips et al. (2006) and Scholl (2007).

<sup>19</sup> I set aside here the issues concerning how things persist, e.g., via perdurance, endurance, and exdurance (cf. Lewis 1986). Debates over these alternative construals of persistence would take us far afield, and are therefore set aside.

<sup>20</sup> Although the research in question has focused mainly on the visual modality, the phenomena in question are not restricted to that modality. There are parallel findings for audition, for example.

<sup>21</sup> For a dissenting view about this interpretation, however, see one of the main investigators in the infancy research, Fei Xu (1997). She claims that "physical object" is itself a sortal. This is a very non-standard view, however, among psychologists (and probably philosophers as well).

<sup>22</sup> My exposition of the two-systems persistence judgment literature draws extensively from the account of Carey (2009). Other summaries of roughly the same research can be found in Rosenberg and Carey (2009) and Scholl (2002).

<sup>23</sup> I am indebted to Holly Smith for the formulation and endorsement of this response and for the gist of how it might be defended.

<sup>24</sup> Thanks to members of my spring 2013 metaphysics seminar at Rutgers University and members of a fall 2013 seminar at Princeton University as well as Jonathan Schaffer, Holly Smith, and Karen Bennett for valuable comments on various drafts of this paper.