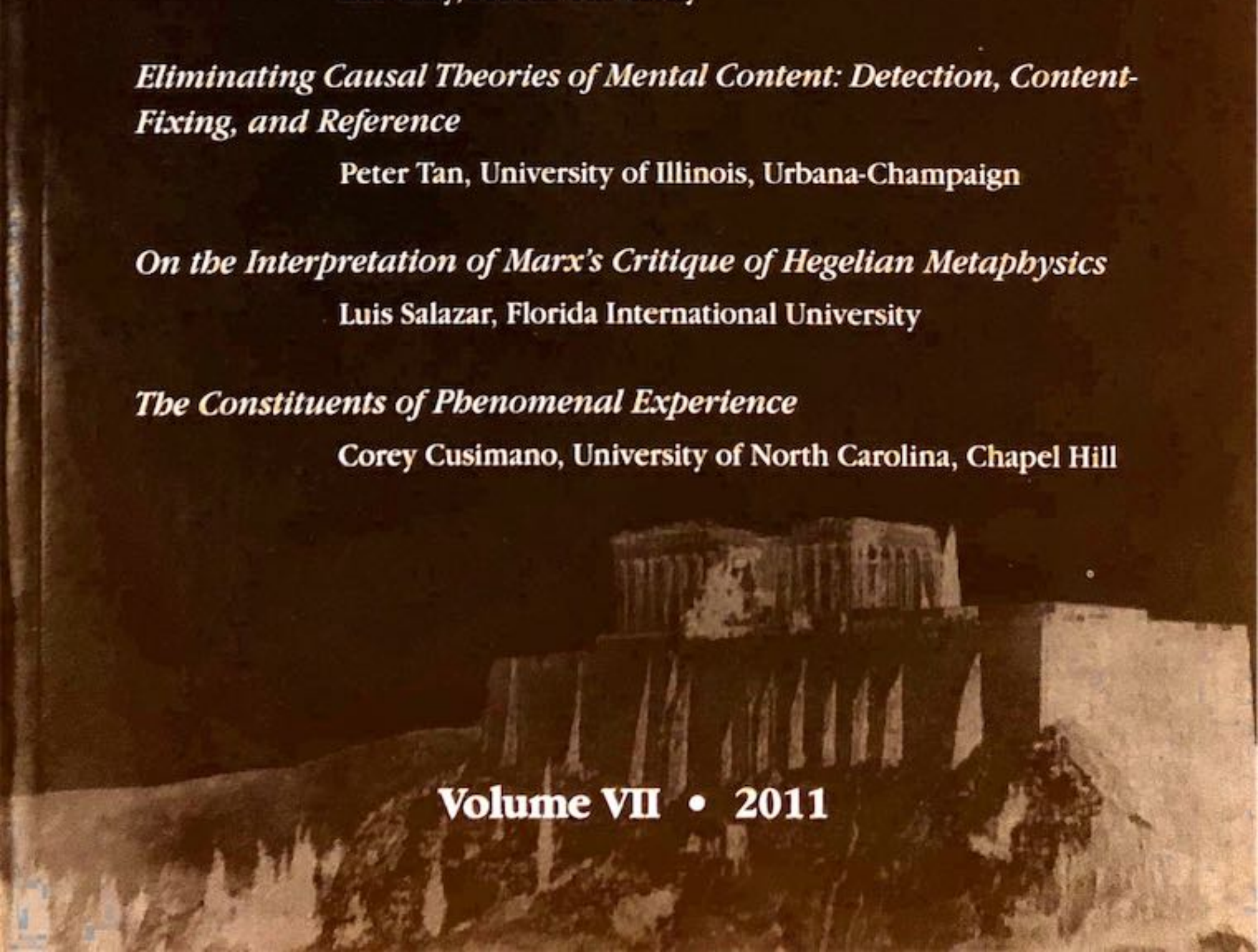
The Cornell Undergraduate Journal of Philosophy

Logos

λόγος



Volume VII • 2011

Logos : λόγος

Volume VII • 2011

An independent student publication

Logos: The Undergraduate Journal of Philosophy, an independent student organization located at Cornell University, produced and is responsible for the content of this publication. This publication was not reviewed or approved by, nor does it necessarily express or reflect the policies or opinions of, Cornell University or its designated representatives.

Mail:
Logos: The Undergraduate Journal of Philosophy
c/o The Sage School of Philosophy
218 Goldwin Smith Hall
Cornell University
Ithaca, NY 14853

Email:
logos-mailbox@cornell.edu

Logos : λόγος

Volume VII • 2011

Contents

Editors' Introduction

It is with great pleasure that we present our readers with the seventh volume of *Logos: The Cornell University Undergraduate Journal of Philosophy* for the 2010-11 academic year. While our primary aim as a peer-reviewed undergraduate publication is to seek out and publish some of the foremost undergraduate philosophy papers available, *Logos* additionally looks to promote the enjoyment of philosophy to the broader undergraduate community and beyond.

In keeping with this goal, *Logos* has had an especially eventful year: Professor Ned Block (NYU) delivered the Norman Kretzmann Undergraduate Philosophy Lecture, in which he spoke about the perceptual phenomena known as "change blindness" and "inattentional blindness." Professor Christine Korsgaard (Harvard) also gave a thought-provoking lecture in October of last year on "Valuing Our Humanity," discussing how animals fit into the scheme of Kantian ethics. We would like to extend our utmost thanks to both of these speakers for their willingness to support undergraduate philosophy at Cornell.

Besides such lectures, discussion groups, and other informal talks hosted by *Logos* throughout the year, we were fortunate enough to hold our first-ever "Life Raft Debate" in late February. This event, originally held by the philosophy department at the University of Montevallo in Alabama, presents its audience with a hypothetical scenario: In the wake of an apocalyptic event, a group of survivors floating on a life raft with only one seat left chance upon a group of professors, each from a different discipline. As they set off to start a new world, they must decide which professor should receive this last spot on the raft. Before voting begins, each professor must address the audience (or "raftees") with what amounts to a spirited (and often humorous!) defense of their academic discipline. We particularly enjoyed this event, as it served to shed light upon the value of very disparate disciplines, while also showing the usefulness and commonality of each subject. The debate's success must ultimately be attributed to the professors who so graciously set themselves at the mercy of the crowd: Professor Karen Bennett (Philosophy); Professor Melanie Dreyer-Lude (Theater, Film, and Dance); Professor André LeClair (Physics); Professor Masha Raskolnikov (English); and Professor Antonia Ruppel (Classics).

Due to a record number of submissions on philosophy of mind, three out of the five articles presented in this volume cover topics in that subject. They range from a thoughtful discussion about what constitutes phenomenal

experience, to a close look at the varying theories of mental content. We hope that the gentle reader will find the focus of these essays both enlightening and thought provoking in how they present differing, but very important aspects of this increasingly popular sub-discipline of philosophy. And while the emphasis on this sub-discipline certainly demonstrates a current trend in contemporary analytic philosophy, it speaks to the diversity of our submissions on the whole that we are able to include an article in the continental tradition. The staff of *Logos* would like to thank all of those who submitted articles for consideration.

As always, without the ever-constant support and funding of the Sage School of Philosophy and the Student Assembly Finance Commission, we would never have been able to bring this journal to fruition. We also extend our utmost gratitude to our undergraduate staff for their hard work and dedication. In closing, we would like to express our heartfelt thanks to our advisor, Professor Michelle Kosch, whose unwavering guidance and support has proved invaluable.

Ariana Marmora
Editor-in-Chief

Daniel Ranweiler
Assistant Editor

Cornell University, Spring 2011

Properties and Dispositions:

On Causal Relevance and Philosophy of Mind

Joanna M. Klimaski
University of Scranton

> *We think we have knowledge of a thing only when we have grasped its cause.* - Aristotle (Posterior Analytics, 71 b 9-11)

Contemporary trends in philosophy of mind have galvanized non-reductive physicalism, the thesis that (1) the world and its components are essentially physical, and (2) entities cannot be reduced to their fundamental physical parts. The hallmark of this theory is its claim that reality is comprised of layers, each one metaphysically affixed to its neighbors while still retaining its own unique ontological status. Higher-level phenomena are thought to be dependent on, but not reducible to, the lower levels.

Ostensibly, a levels ontology resolves the problems bequeathed by a Cartesian worldview as well as those that come with radical physicalism. But ultimately we face the same questions that plague these views. How can disparate substances interact? How does the physical cause the mental, and vice versa? Moreover, can purportedly higher-level phenomena, such as beliefs, cause lower-level phenomena, namely neural reactions, as well as other higher-level phenomena, such as other beliefs? There must be some causal nexus or an ontological junction at which the realms meet such that they produce effects in one another.[1] Without this point whereby the causal relation occurs, non-reductive physicalism, "like Cartesianism, founders on the rocks of mental causation."[2]

John Heil addresses these questions in his recent work *From an Ontological Point of View*. It is evident from this text than an adequate theory of causality, particularly of mental causation, must stem from an adequate theory of properties, something current non-reductionist theories overlook. Moreover, by proposing a revisionary ontology of properties, Heil also offers a promising account of causality that avoids the common drawbacks of completeness, exclusion, and causal overdetermination.

In this paper I present common problems of mental causation via three prevalent non-reductionist theories: property dualism; functionalism; and emergentism. I attempt to show how Heil's ontology of properties can resolve some of the issues these theories encounter with regard to mental causation. Additionally, I illuminate relevant philosophical issues such as free will and determinism. In applying Heil's insights about the nature of properties, I argue that it becomes apparent that these so-called problems of mental causation are in fact not problems at all.

---

[1] "Mental Causation," *The Stanford Encyclopedia of Philosophy*, accessed November 28, 2009 <http://plato.stanford.edu/entries/mental-causation/> §2.1.

[2] Jaegwon Kim, "The Non-Reductivist's Troubles with Mental Causation," in *Mental Causation*, ed. John Heil et al. (Oxford: Clarendon Press, 1993), 193.

I. THREE THEORIES OF MENTAL PROPERTIES

Non-reductive physicalist theories affirm that the mind is a physical entity, but deny that as such it can be reduced to its fundamental physical parts. Hence, non-reductionists grant a distinction between "physical properties" and "mental properties." Mental properties are dependent upon, but irreducible to, and for this reason ontologically distinct from, physical properties. Though pain involves a series of underlying physical activities, the experience of being in pain is something over and beyond them. Jaegwon Kim describes non-reductive physicalism as such:

> Its ontology is physical monism, the thesis that physical entities and their mereological aggregates are all that there is; but its 'ideology' is anti-reductionist and dualist, consisting in the claim that psychological properties are irreducibly distinct from the underlying physical and biological properties... [but] genuine properties nonetheless, as real as underlying physical-biological properties.[3]

Among the more popular non-reductive theories are property dualism, functionalism, and emergentism. I will deal with each of these separately.

Property dualism is an attempt to acknowledge the apparent dissimilarity between physical and mental *properties*, without propounding two distinct *substances*, as is characteristic of Cartesian dualism. Thus, on the property dualist's view, all *substances* are essentially physical, but some of them have two kinds of metaphysically distinct properties: mental properties and physical properties. Mental properties are not reducible to physical properties, but are something over and above the physical. These properties are irreducible and ontologically unique, though not separate substances, independent of the physical entities which possess them. Rather, they are features of the composite (physical) entity which possesses them.[4]

Someone familiar with the pitfalls of Cartesian dualism will immediately see a classic problem: how could ontologically distinct properties interact causally? This question signifies what is called the completeness problem, which appeals to the causal closure of the physical universe. As modern science would tell us, the physical world is "causally closed."[5] This means that exceptionless physical laws govern the world; hence, all physical events have completely

---

[3] Kim, "The Non-Reductivist's Troubles with Mental Causation," 192.

[4] John R. Searle, "Why I Am Not a Property Dualist," *Journal of Consciousness Studies* 9 (2002): 57-64.

[5] John Heil, *From an Ontological Point of View* (Oxford: Clarendon Press, 2003), 20.

physical explanations.* If this is the case, all things mental are epiphenomenal, or secondary to physical events. Taking epiphenomenalism a step further, one can say that the mental plays no role whatsoever if all events are in fact explainable by the physical. Thus, the completeness problem for property dualism entails that at best mental properties are derivative, and at worst they are fictitious. Physical effects must have, at least in part, physical causes.[6] Otherwise there are effects *ex nihilo*.

Another prevalent theory of mind is functionalism. As its name suggests, functionalism underscores the functional roles of mental states. Though functionalism shares the basic tenet of non-reductive physicalism, which states that mental properties are distinct from physical properties, functionalists do not claim (unlike property dualists) that mental properties are autonomous. On this account, mental properties are functional properties.[7] That is to say, a mental property is that which causally mediates the physical source of an event, a sensory stimulus, and a behavioral effect (or another mental state). The mental state "being in pain," for instance, is what causally mediates the event of "Sam the golden retriever biting me" and my subsequent shrieking.

A convenient upshot of functionalism is multiple realizability, the idea that ostensibly similar mental properties can manifest in various species, despite disparate physical structures and make-up. For instance, identifying the mental state of "being in pain" with a certain neurological process in human brains is problematic insofar as we attribute the sensation of "being in pain" to creatures with vastly dissimilar neurological structures from ours. Is this possible, given the diversity of neurophysiology? Can we meaningfully and non-equivocally say that a dolphin is "in pain" in the same way that we say a human is "in pain?" To account for this, functionalists propose that diverse entities can all be said to "be in pain," because each undergoes a causal process that connects a sensory input (that would cause pain) with the behavior exhibited as an effect. Whether the sensory inputs or the ensuing behavioral outputs are similar across species is irrelevant; at a minimum, the causal process is analogous.

However, as with property dualism, functionalism fails to elude certain obstacles associated with mental causation. Here we encounter the exclusion problem. Exclusion is another face of epiphenomenalism, in that it queries the relevance of mental properties in the overall causal scheme.[8] By granting mental properties ontological import, yet conceding their dependence on lower-level physical properties, the functionalist cheapens the causal contribution of mental properties. If the causal efficacy of mental properties ultimately comes down to

---

* Whether we yet possess a complete explanation is a separate, epistemological issue.

6 "Mental Causation," §2.3.

7 "Mental Causation," §6.1.

8 "Mental Causation," §6.2.

the causal authority of their physical partners, what role does the mental really play? Its own contributions would seem to be pre-empted by the physical.

By leaving us for want of a robust understanding of causation, functionalism "answers the metaphysical question without answering the ontological question."[9] It answers the question of how mental properties work and how they apply across diverse physical structures, but it fails to illuminate the issue of what mental properties *are*. If they are not physical, presumably they are non-physical, and thus we are in the same predicament as we were with property dualism. If mental properties do in fact causally interact with physical properties, they must share some characteristics (perhaps physical ones) with the physical. Do mental properties have both physical and non-physical parts? Functionalism is silent on these issues.

Finally, emergentism is the thesis that mental properties are emergent properties, i.e., that they are ontologically distinct, higher-level properties that emerge from highly complex physical systems. On this account, mental phenomena are the result of exceptionally complex brains which give rise to conscious experience. Jaegwon Kim explains:

> ...the intuitive idea of an emergent property stems from the thought that a purely physical system, composed exclusively of bits of matter, when it reaches a certain degree of complexity in its structural organization, can begin to exhibit genuinely novel properties not possessed by its simpler constituents.[10]

Beliefs, desires, intentionality, and emotional affect, then, are said to emerge from neuronal substrata. Moreover, all mental phenomena are determined by configurations of neural events, such that "if the very same configuration of physiological events were to recur, the same mental phenomenon... would emerge again."[11] The relationship between mental properties and their physical underpinnings, then, is one of necessitation.

One important component of emergentism is the supervenience thesis.

> *Supervenience*: If property M emerges from properties N_1..., N_n, then M supervenes on N_1...,N_n. That is to say, systems that are alike in respect of basal conditions, N_1...,N_n must be alike in respect of their emergent properties.[12]

---

[9] Samuel Guttenplan, *A Companion to the Philosophy of Mind* (Oxford: Blackwell Publishers Ltd., 1995), 326.
[10] Jaegwon Kim, "Emergence: Core ideas and issues," *Synthese* 151 (2006): 548.
[11] Ibid., 550.
[12] Ibid.

In less formal terminology, the supervenience thesis proposes that mental properties supervene on physical properties such that there could be no change in the supervenient mental properties without likewise effecting a change in the underlying physical properties. Despite this intimate connection between emergent properties and their basal properties, emergent properties are irreducible and thus cannot be explained in terms of physical properties.[13]

Irreducibility and inexplicability introduce the first problem that plagues emergentism. By appealing to the primitive nature of emergent properties (saying that they cannot be analyzed or reduced), we preclude any hope of shedding ontological light on them. That is, we are given no account of what they *are*. Though we want to know what it is that grounds the supervenience claim (i.e., by virtue of what do supervenience and emergence occur), it seems we must accept their inexplicability.[14]

More germane, emergentism backfires with regard to mental causation. Like all non-reductive physicalists, Kim explains, emergentists are realists about mental properties.

> To be real, arguably, is to have causal powers. Anything real must be part of the causal structure of the world. So if mental properties are real features of the world, they must have causal powers; that is, having a mental property must endow the thing that has it with powers to affect courses of events in its neighborhood.[15]

All properties, physical and mental alike, have causal powers; for, properties that do not contribute causally to their possessors contribute nothing at all. Thus, emergent properties must have distinctive causal powers which are irreducible to the causal powers of their basal properties.[16] To account for this, emergentists invoke the presence of causal laws. We know that the laws of nature account for lower-level physical events. The law of gravity, for instance, assures that objects fall at an average rate of 9.8 m/s^2 on Earth. We can go further by positing additional laws that govern higher-level events in the same way that lower-level laws do. These laws are neither reducible to, nor derivable from, the laws of nature.[17]

Causally efficacious mental properties and a subsequent theory of level-specific laws raise the problem of causal overdetermination, the claim that

---

[13] Kim, "Emergence: Core ideas and issues," 551.
[14] Ibid., 556.
[15] Jaegwon Kim, *Philosophy of Mind* (Boulder: Westview Press, Inc., 1996), 230.
[16] Kim, "Emergence: Core ideas and issues," 557.
[17] Heil, *From an Ontological Point of View*, 32.

events cannot have more than one sufficient cause. Say that the physical state of dehydration, P_1, causes my putative mental desire for water, M_1. My desire for water then causes my intention to go to the refrigerator, M_2. But M_2 is caused by its own physical phenomenon, P_2. Now M_2 seems to have two sufficient causes, M_1 and P_2, and thus is causally overdetermined. If we are inclined to circumvent this problem by saying that P_2 does not exist and that M_2 is accounted for by M_1 alone, then we violate the principle of causal closure of the physical world because M_2 would lack a physical explanation.

Though these ontologies indeed have merit regarding various technical problems in philosophy of mind, all of them fail when it comes to mental causation. Such pitfalls, in all of these theories, are symptomatic of a deeper problem. Their flawed notions of causation are rooted in mistaken concepts of the nature of *properties*. Despite an alleged departure from Cartesian dualism, each theory nonetheless maintains a distinction between the mental and the physical. Property dualism posits this distinction outright, while functionalism and emergentism grant it in their implicit allegiance to a *levels ontology*. Such a view holds that reality is comprised of various levels, each ontologically distinct from and thus irreducible to one another, yet all of them dependent on their neighboring levels. *Prima facie*, it is an appealing theory, as it allows us to differentiate between such ostensibly dissimilar properties as having a belief and being an axon terminal. Inevitably, however, all theories that are anchored in this idea run into the various problems of causation which I have discussed. They splinter the world into categories and cannot get the theoretical pieces back together again. And so we need a new ontology, one that restores cohesion among properties.

II. HEIL'S ONTOLOGY OF PROPERTIES

Often, contemporary theories of mind do not address the issue of what properties *are*, but rather presuppose a faulty ontology. Many maintain an implicit allegiance to the "picture theory of meaning," or the idea that the character of reality can be ascertained from our linguistic representations of reality.[18] This generates the notion that each predicate literally corresponds to the property it represents. Ontologies, such as the bundle theory of properties, which holds that objects are composites of properties, emerge from this. Think of objects on this view as Mr. Potato Head. Just as Mr. Potato Head is fundamentally a potato which possesses a collection of eyes, ears, mustaches, and accessories, so objects are fundamental substances to which properties adhere. Without the

[18] Heil, *From an Ontological Point of View*, 6.

metaphysical glue which could serve to assemble properties, the better move is to proffer an ontologically serious view of properties.

In his revisionary work *From an Ontological Point of View*, John Heil proposes that properties are intrinsic powers or "dispositionalities."* Properties are not aspects or parts of an object, but *ways* of an object. Put differently, they make objects the way they are by virtue of their standing in relation to one another at certain instances in time. Objects are property-bearers, but not bundles of properties. A shirt, for instance, is not merely a bundle of threads. Rather, it occurs from the distinctive relation that the threads bear to one another during a given stretch of time, during which the threads dispose the shirt to be a certain size, shape, color, to be appealing to its wearer, etc.

To be real, Heil says, is to possess causal powers.[19] Properties are certainly not exempt from this dictum. A causally inert property would seem to make no difference at all to its possessor.[20] The distinctive feature of properties, then, is their inherent capacity, or power, to dispose their possessors to behave in certain ways or to cause certain effects in conscious observers. They make distinctive causal contributions to their possessors.[21] Properties are not, however, *pure* powers, that is, exclusively dispositional. When we speak of properties, we must also speak of their qualitativity. Qualitativity refers to the intrinsic, categorical qualities an object possesses, for instance its color or its shape. Dispositionality and qualitativity, then, are the two characterizing facets of properties.

The distinction between these facets, however, goes no further than characterization. They are not *ontologically* distinct. In fact, Heil contends that a property's dispositionality and qualitativity are one and the same. This view is the "identity theory of properties":

> (*IT*) If P is an intrinsic property of a concrete object, P is simultaneously dispositional and qualitative; P's dispositionality and qualitativity are not aspects or properties of P; P's dispositionality, P_d is P's qualitativity, P_q, and each of these is P: $P_d = P_q = P$.[22]

Properties are simultaneously qualitative and dispositional: qualities are dispositional and dispositions are qualitative. For instance, a baseball's sphericity is both an inherent quality of a baseball and a disposition to roll.

It is in divorcing dispositionality and qualitativity that theories of properties,

---

* Heil uses "powers" and "dispositionalities" interchangeably.

[19] Heil, *From an Ontological Point of View*, 97.

[20] Ibid., 77.

[21] Ibid., 76.

[22] Ibid., 111.

and ultimately theories of mind, founder. Allowing this fundamental distinction countenances a world of levels in which dispositionality is grounded in, or supervenes on, non-dispositional, categorical (qualitative) properties. This view accordingly brings us back to the primary problem of emergentism: by virtue of what does supervenience occur? Though dispositionality and qualitativity can be separated conceptually for explanatory purposes, levels of explanation do not translate to levels of reality without ensuing metaphysical issues.

Many theorists subjugate the behavior of properties to the authority of unconditional laws of nature. The instantiation of properties and their subsequent behavior is contingent upon the presence of these laws. Hence, if an object's qualitative properties were combined with different laws of nature, the object would have different dispositionalities than those it had with prior laws.* Heil, however, inverts the relationship between laws and properties. All laws, he says, are themselves grounded in powers possessed by objects. Laws are what they are by virtue of an object's properties, i.e., an object's ways. These ways *cause* objects to behave in certain law-like regularities under certain conditions. Laws make known the causal powers of objects by standing mediately between the properties themselves, and the effects that properties produce in either other objects or conscious observers. Thus, on Heil's view, causality is grounded in laws, which in turn are grounded in the dispositions of objects.[23]

Causation is not a numinous interaction between ontologically distinct levels, as property dualists and emergentists maintain, nor is it a bridge between the levels, as functionalists contend. Rather, it is grounded in an object's properties *qua* powers *qua* qualities. By virtue of possessing certain properties standing in relation to one another, objects are disposed to behave in certain ways or to effect certain experiences in conscious observers.

With this more robust account of properties, we are in a better position to address mental causation. Properties, we said, are intrinsic dispositionalities/ qualities of objects. Properties are particularized *ways* objects are. Mental and physical properties, with regard to conscious agents, are intrinsic powers that make distinctive causal contributions to their possessors. But just as Heil contends that distinctions between dispositionality and qualitativity are for the purpose of characterization only, and are not ontological divisions, I propose that distinctions between the physical and the mental serve similar purposes. Talk of physical properties and mental properties are conceptual tools by which we characterize the world and experience. More importantly, I believe "physical" here is a misnomer, which leads to additional confusion regarding this distinction. I will return to this point shortly.

---

* D.M. Armstrong is a proponent of this view. See Heil, pp. 120-121.

[23] Heil, *From an Ontological Point of View*, 66.

Although such compartmentalization is useful for explanatory purposes, our conceptual and linguistic distinctions should not lead us to conceive of reality as stacked with myriad levels threaded together by some mysterious metaphysical stitching. There are not mental properties and physical properties, but only properties: qualitative/dispositional powers. Having height, mass, organs, nervous systems, consciousness, and beliefs are all *ways* familiar sorts of conscious beings are. All of the ways make distinctive causal contributions to their possessors. Having a certain mass disposes me to makes footprints in the mud. Having a highly complex nervous system disposes me to be conscious. Having beliefs about closet monsters disposes me to demonstrate fearful behavior. For conscious agents in particular, brains have intricate neurophysiological structures which dispose their hosts to behave in particular ways and to have experiences.

It is important to mention how this view escapes the threat of reductionism, the view that complex entities, such as consciousness, can be reduced to their constitutive parts (typically neurons and microphysical reactions in the brain). Reductionist tendencies reveal yet another linguistic mirage conceived of by modern philosophers, that is, that language designates literal entities. Heil locates this widespread tendency in an implicit devotion to the picture theory of meaning, which presupposes a one-to-one relation between predicates and their corresponding referents. Endorsing this view would lead us to think that the phrases "being in pain" or "having a belief" refer to precise properties, and conversely, that the state of pain that I am in now is exactly what I mean when I say that "I am in pain."

Heil argues instead for imperfect similarity. Pains and beliefs can take many forms which manifest similarly enough to warrant ascribing the term "pain" or "belief," but do so without being *perfectly* similar. Thus our predicates, imperfectly attempting to classify imperfectly similar events, pick out what may seem to be one occurrence or one entity, such as being in pain, but what is in fact a multiplicity. Heil's view, therefore, is not reductionism because Heil makes no move to reduce predicates to particular properties, or properties to particular neural configurations. A predicate denotes various kinds of properties, or even various combinations of properties. Likewise, a property need not refer to one particular entity.

III. HEIL AND MENTAL CAUSATION

Through the three theories of mind discussed above, I have presented three corresponding problems of mental causation, viz., the completeness problem, the exclusion problem, and causal overdetermination. The true test of Heil's ontology, then, is its ability to withstand these problems. As above, I will

contend with each of these problems separately.

Recall that according to the principle of causal closure of the physical world, all physical events are explainable exclusively by their physical underpinnings. This principle, from which stems the completeness problem, does not present a difficulty for Heil's view of properties as powers since this view, at its core, is consonant with physicalism. Properties are essentially physical, that is, grounded in the physical universe. Further qualifying properties as either physical or mental proposes a false dilemma, for what the physical/mental distinction actually implies is a distinction between *material* and *immaterial*. On the face of it, this would indeed leave us with the problems that arise in the so-called physical/mental distinction. Note, however, that it is completely natural to posit both material and immaterial entities as commonplace *physical* phenomena. Gravity, for instance, is an *immaterial* but nonetheless *physical* force. Properties, whether material or immaterial, belong to physical reality, and as such do not violate the principle of causal closure.

The exclusion problem queries the causal relevance of mental properties. If mental properties ultimately depend on their physical realizers, then their potential causal contribution is already present in their underlying realizers. In other words, the causal powers of mental properties are otiose given the causal powers of physical properties. This again presupposes an ontological distinction between physical and mental. By letting go of a levels ontology, vestiges of a dualistic worldview, we become less inclined to regard the mind as comprising "physical" and "mental" parts. Rather, there are just properties. Our properties *qua* conscious agents dispose us to think, behave, and desire; simultaneously, they dispose us to breathe, digest, and circulate blood. Jettisoning levels among the physical and the mental expunges claims of the causal ineffectiveness of numinous secondary properties.

Hence, as we have seen so far, a Heilian view of properties eliminates common problems regarding mental causation. The problem of causal overdetermination, however, will require more explaining. Recall that in the case of effect *E* having two or more sufficient and distinct causes *X* and *Y*, *E* is overdetermined.[24] If my desire for water, M_1, causes an intention to go to the refrigerator, M_2, and my intention is caused by its own physical groundwork, P_2, then the intention is overdetermined: M_1 and P_2 are both sufficient to cause M_2. Put another way, "independent overdeterminers can 'come apart' – that is, either one of those causes could occur without the other."[25] And, if they are sufficient conditions, the effect, M_2, would still occur.

Applying Heil's insights about the nature of imperfect similarity, however,

---

[24] Eric Funkhouser, "Three Varieties of Causal Overdetermination," *Pacific Philosophical Quarterly* 83 (2002): 335.
[25] Ibid., 342.

makes this last statement is inaccurate. M_1 and P_2 cannot in fact *come apart* because they are not discrete causes. In order to show this, I will briefly reiterate Heil's conception of properties and causation. Properties, which are simultaneously qualitative and dispositional, are ways of objects. Thus they dispose objects to behave in certain manners and to have certain effects on conscious observers. They make objects what they are. Now translate this conception of properties and objects to properties and conscious agents. Conscious agents have certain properties standing in relation to each other at certain times, all of which make their possessors the way they are – here, living organisms that are conscious. Properties are not distributed among a multilayered hierarchy of being that consists of physical properties and mental properties, because the existence of ontologically distinct levels of reality would preclude any interaction among the levels. It follows, then, that there is one level of reality, and it is to this level that properties, objects, and conscious agents belong.*

It is properties which make causation possible. Causation, on Heil's view, is grounded in laws, which are in turn grounded in properties, viz., the powers objects possess. Laws are indicative of the ways objects are disposed to behave under certain conditions. Causation, transitively, is grounded in these dispositions. Recall Heil's contention that predicates do not correspond literally to properties. For instance, when I note that the apple in my hand is red, the predicate "red" does not identify a universal property. The red of the apple is different from the red of a cardinal, yet I can correctly describe both as red. Thus it is our language that bewitches us into averring a uniform relation between descriptive predicates and the properties they designate. Just as there is no one-to-one relation between predicates and properties, I argue that no one-to-one relation exists between dispositions and effects, that is, between properties and the ways in which they manifest themselves. Saying that "M_1 causes M_2" refers to a particular causal relationship, is symptomatic of the mistaken theory of properties which Heil warns about (one which assumes that the properties we name single out particular entities in reality). Heil instead contends that properties are *ways* standing in relation to one another. This does not indicate that ways are particular entities or particular manifestations.

What bearing does this have on causal overdetermination? The absence of a one-to-one relationship between particular dispositions and particular manifestations would also suggest the absence of one-to-one relationships between causes and effects. This would mean that for many of the ordinary causal relations that we speak about, *there is no one sufficient condition*. Just as no one property entails a particular manifestation, there is no one condition that

* Heil posits this substance as the quantum field, or space-time. Though the nature of this one substance is itself a fascinating topic, it is not within the scope of this paper.

entails a particular event. Instead, there is a set of several necessary conditions (conditions *in whose absence* an event *E cannot* occur) that together comprise the sufficient condition (the condition *in the presence of which* the event *E must* occur).[26] Our language would lead us to believe that what we refer to as a sufficient cause is the single cause, when in fact there are likely several causes, conditions, properties, or events standing in relation to one another. These contribute gradationally, and some are more or less important than others (for instance, remote or proximate causes).

J. L. Mackie refers to the conditions that I am describing as INUS conditions, an acronym which describes a cause that is "an *Insufficient* but *Necessary* part of a condition which is itself *Unnecessary* but *Sufficient* for the result."[27] According to Mackie, our common notions of causation, for instance, that *A* caused *P*, isolate one cause, *A*, as the sole determiner of an event. Yet these notions overlook a broader recipe of causes, *B*...*n*, which likewise facilitated the event. *A*, then, is not the sufficient cause, because it could not have produced *P* independently. But it is not a necessary cause either, since, in its absence, another cause, *J*, could have perhaps substituted for it. Instead, *A* belongs to a set of factors (both positive and negative) which conjunctively form a complex condition, *X* (*A*, *B*...*n*). Like *A*, *X* is not a necessary condition, because another well-suited combination of causes may as well have brought about the event. *X* does, however, constitute the minimal sufficient condition for the event.

Consider a car accident. On the face of it, it may seem that the large Volvo that veered out of its lane was the sole cause of my Honda's crashing into the guard rail. But in fact, this ill-fated vehicle was merely the proximate cause, or, in Mackie's terms, the INUS condition. The Volvo is not a necessary cause, since a Toyota in a similar situation could very well have brought about this event. And alone, a veering Volvo does not itself ensure a crash. If the driver of the Volvo had slept a reasonable amount the previous night, and so had not fallen asleep at the wheel, or if I had not averted my eyes to speak to my passenger, or if recent rainfall had not slicked the roads, then the event of hitting the guard rail could have been prevented. Yet the presence (or absence) of other conditions, such as the rainfall and the lack of sleep, combined with the most patent cause, the Volvo, are all factors of the complex condition that in fact led to catastrophe.

One might object that, in taking into account a multiplicity of causes for an event, virtually anything can be included. For instance, my birth is technically a necessary condition for my crashing into the guardrail: if I had not been born, I would not have been in a car accident. And since what we assert in causal statements pertains to some fact or set of facts about the world, we can ultimately

---

[26] Carl Cohen and Irving M. Copi, *Introduction to Logic* (Upper Saddle River: Pearson Prentice Hall, 2005), 470.
[27] J. L. Mackie, "Causes and Conditions," in *Metaphysics: An Anthology*, ed. Jaegwon Kim et al. (Boulder: Blackwell Publishing, 1999), 415.

consider the whole world as a participant in this causal relation.

It seems excessive, however, to include my entire life and even the entire universe in this one event. Thus, to delimit the relevant causes for a given event, Mackie introduces the notion of a causal field, a "wider region in which the effect sometimes occurs and sometimes does not."[28] In other words, the causal field is the abstract domain in which the causal factors and conditions, as well as the unrelated features, occur. Though the actual boundaries of a causal field are only approximate, they serve to provide a backdrop to the causal relation and conceptually limit the plethora of possible factors. For instance, if we were seeking to determine the cause of diabetes in dogs, we would limit our causal field to *the class of dogs in general*, rather than to, say, every creature that can be affected by diabetes. In the singular case of my car accident, we can designate the causal field as *the class of Hondas that crash into guardrails*. It includes all the features that were implicated in the event (the Volvo; sleep-deprivation; rainfall) as well as other features that were present, but not in fact implicated in the crash (Route 78; the steel of the guardrail; the hitchhiker who provided a testimony of the accident).

Thus the matter of causal overdetermination can be summarized as follows. Overdetermination among the physical and the mental does not occur, because both "physical" and "mental" refer to the same level of reality, said in different ways. Moreover, even without levels, there may be no one sufficient cause for any given event, but rather a complex condition which comprises several causes, some remote and others proximate. What we call property *X* does not refer to the same entity in every instance. We perceive imperfectly similar manifestations of various ways objects are (viz., properties), and we apply our limited vocabulary and conceptual repertoire in an attempt to categorize what we perceive. The redness of an apple is not the same redness of the apple two days later, and neither of these is the same redness of a fire truck. The tendency to oversimplify the associations between predicate and property, and between property and manifestation, likewise translates to the associations we make between cause and effect. By naming a sufficient cause *X* for a given effect *P*, we identify *X* as the causal partner of *P*. But *X* may in fact refer to a complex condition consisting of several causes standing in relation to each other at particular times, all of which yield *P*.

IV. CAUSAL OVERDETERMINATION AND DETERMINISM

In addition to furthering understanding about the nature of the mind and of conscious agents in general, the above conclusions have interesting implications

[28] Mackie, "Causes and Conditions," 416.

for the metaphysics of free will. This classic debate examines the question of whether human beings have freedom of will or whether they are determined by preceding events that leave them without choice in their ostensible decisions. The classic formulation of the free will problem goes as such:

> 1. Some person (*qua* agent), at some time, could have acted otherwise than she did.
> 2. Actions are events.
> 3. Every event has a cause.
> 4. If an event is caused, then it is causally determined.
> 5. If an event is an act that is causally determined, then the agent of the act could not have acted otherwise than in the way that she did.[29]

If all events have a sufficient cause, which by definition is a cause in the presence of which the event *must* occur, then given the occurrence of certain causes, an agent, contrary to what she might think about herself and her agency, is without choice regarding the ensuing event. Her actions are wholly determined by the set of conditions in which she finds herself.

Peter Van Inwagen illustrates the free will problem as a garden of forking paths. The main path which forks into several diverging paths represents a decision that an agent must make and the various options and outcomes she has as a result.

> To say that one has free will is to say that when one decides among forks in the road of time (or, more prosaically, when one decides what to do), one is at least sometimes able to take more than one of the forks... One has free will if sometimes more than one of the forks in the road of time is "open" to one. One lacks free will if on every occasion in which one must make a decision only one of the forks before one – of course it will be the fork one in fact takes – is open to one.[30]

Determinists, on the other hand, hold that there is no such thing as free will.

> Determinism is the thesis that it is true at every moment that the way things then are determines a unique future, that

---

[29] "Compatibilism," *The Stanford Encyclopedia of Philosophy*, accessed November 30, 2009, <http://plato.stanford.edu/entries/compatibilism/> §1.5.

[30] Peter Van Inwagen, *Metaphysics* (Boulder: Westview Press, 2002), 202-203.

> only one of the alternative futures that may exist relative to a given moment is a physically possible continuation of the state of things at that moment.[31]

To exemplify this point, Van Inwagen describes a situation in which God rolls back history and then lets things play forward again.[32] If determinism is true, then the exact course of events would recur as they had the first time, because each cause would reproduce the effect that had occurred originally. If determinism is false, then an entirely new set of events would likely occur, given the myriad of options each cause entailed. This line of reasoning should sound familiar. On the emergentist view, all mental phenomena are determined by configurations of neural events, such that "if the very same configuration of physiological events were to recur, the same mental phenomenon... would emerge again."[33] A particular configuration of neural events is sufficient for the emergence of a certain mental property. Such a view, as we have seen, ultimately leads to the problem of causal overdetermination, since any mental phenomenon, M_2, which is caused by another mental event, M_1, as well as by its underlying physical realizer, P_2, is overdetermined.

A resolution of the problem of causal overdetermination, then, has a bearing on determinism. Eliminating causal overdetermination means that there is no single, sufficient cause for a given effect. Rather, there are various necessary causes standing in relation to each other at particular times that yield an effect. In Van Inwagen's terms, the lack of a sufficient cause (that is, the lack of a one-to-one relation between cause and effect) means that it is not the case that for every occasion where one must make a decision, only one fork is open to her. And saying that there is more than one fork open to her means that determinism is false.

V. CONCLUSION

The paucity of theories which can adequately account for mental causation speciously portrays causation as a mysterious force in the workings of the mind. It would seem that immaterial entities, such as mental phenomena, can have no impact on their physical foundations, or vice versa, given the absolute disparity between the two. This comes from an erroneous view that grants the theoretical physical/mental division too much ontological clout. Following Heil, by viewing causality as grounded in laws, and laws as grounded in the powers

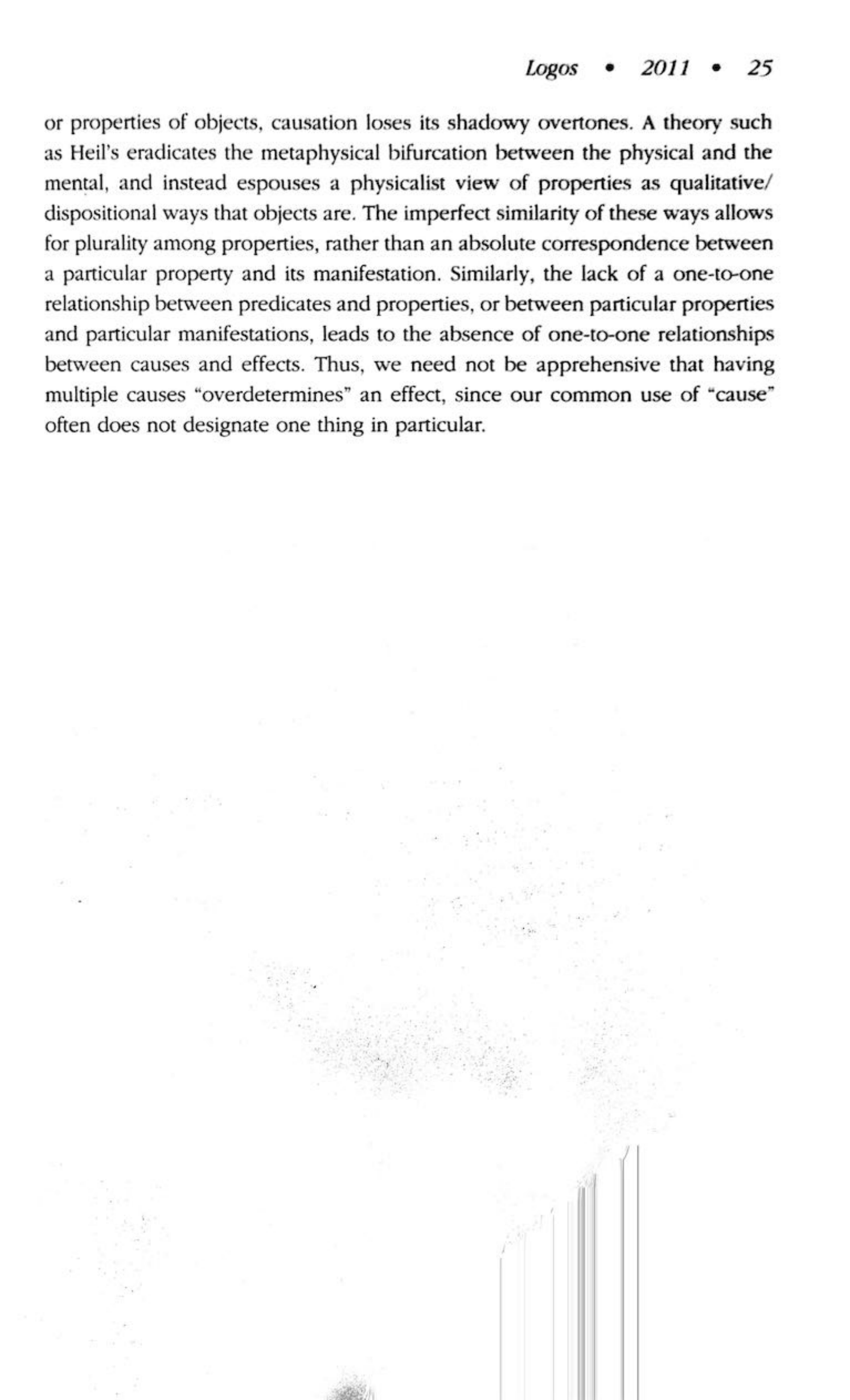
---

31 Van Inwagen, Metaphysics, 203.

32 Ibid., 204.

33 Kim, "Emergence: Core ideas and issues," 550.

or properties of objects, causation loses its shadowy overtones. A theory such as Heil's eradicates the metaphysical bifurcation between the physical and the mental, and instead espouses a physicalist view of properties as qualitative/dispositional ways that objects are. The imperfect similarity of these ways allows for plurality among properties, rather than an absolute correspondence between a particular property and its manifestation. Similarly, the lack of a one-to-one relationship between predicates and properties, or between particular properties and particular manifestations, leads to the absence of one-to-one relationships between causes and effects. Thus, we need not be apprehensive that having multiple causes "overdetermines" an effect, since our common use of "cause" often does not designate one thing in particular.



BIBLIOGRAPHY

Cohen, Carl, and Irving M. Copi. *Introduction to Logic*, 12th Edition. Upper Saddle River: Pearson Prentice Hall, 2005.

Funkhouser, Eric. "Three Varieties of Causal Overdetermination." *Pacific Philosophical Quarterly* 83, No. 4 (2002): 335-351.

Guttenplan, Samuel. *A Companion to the Philosophy of Mind*. Oxford: Blackwell Publishers Ltd., 1995.

Heil, John. *From an Ontological Point of View*. Oxford: Clarendon Press, 2003.

Kim, Jaegwon. "Emergence: Core ideas and issues." *Synthese* 151 (2006): 547-559.

—. *Philosophy of Mind*. Boulder: Westview Press, Inc., 1996.

—. "The Non-Reductivist's Troubles with Mental Causation." In *Mental Causation*, edited by John Heil and Alfred Mele, 189-210. Oxford: Clarendon Press, 1993.

Mackie, J. L. "Causes and Conditions." In *Metaphysics: An Anthology*, edited by Jaegwon Kim and Ernest Sosa, 413-426. Boulder: Blackwell Publishing, 1999.

McKenna, Michael. "Compatibilism." *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. October 5, 2009, http://plato.stanford.edu/entries/compatibilism/.

Robb, David, and John Heil. "Mental Causation." *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. June 27, 2008, http://plato.stanford.edu/entries/mentalcausation/.

Searle, John R. "Why I Am Not a Property Dualist." *Journal of Consciousness Studies* 9, No. 12 (2002): 57-64.

Van Inwagen, Peter. *Metaphysics*. Boulder: Westview Press, 2002.

Narrow Ineffability

Luc Cary
McGill University

Philosophers of music have always considered science important to their field of work.[1] But with modern technological developments, especially in brain imaging, and new methods in psychology, the past thirty years have significantly changed our understanding of music's role in the brain. Scientific studies focusing on music and the brain appeared in large numbers in the 1980s with the emergence of a new discipline: the cognitive science of music. Dr. Diana Raffman is an important player in this field. Her essay, "Music, Philosophy, and Cognitive Science," details three central themes: her theory of musical grammar; her claim that music and language are somehow connected; and the relationship between science and philosophy in studying music. These ideas are developed in her book *Language, Music, and Mind*, in which she discusses her theory of musical ineffability.

Diana Raffman relies on neuroscientific and psychological studies as support for her theory, which allows the comparison of her views with arguments from the philosophy of science. Interestingly enough, her explanation of "M-grammar" nearly mirrors the covering law model for scientific explanations. Also, her reliance on science permits discussions of empiricism and induction. I will, for the most part, analyze a specific claim of hers: Raffman suggests that science can solve a given problem until that problem becomes confounding enough to solve by using philosophy.

Although some examples may follow this construct, I do not think that the problem of "musical ineffability" is such an example. The statement that there is a point at which science ceases to solve problems, and that past this point only philosophy can be the solver, is far too narrow in its scope given such a broad subject. I hope to show this through examples of inductive reasoning, scientific realism, and self-contradiction. This paper aims to critique a largely aesthetic philosophy by utilizing principles from the philosophy of science.

INEFFABILITY 101

In *Language, Music, and Mind*, Raffman maintains that our understanding of nonlinguistic arts is *ineffable*. There are some aspects of artwork that we cannot adequately put into words. The saying "a picture is worth a thousand words" comes to mind when reading her discussion of ineffability, in that humans can tell when certain things are beyond verbal explanation. Raffman describes music, however, as a uniquely ineffable art.

Music, unlike other nonlinguistic arts, has its own grammatical structure. This does not necessarily mean that there is a language of music: even if we understand its grammar, we cannot (yet) communicate through music the way

[1] Diana Raffman, forthcoming in *Routledge Companion to Philosophy of Music*, 2011.

that we can through speech. Consequently, Raffman is not claiming that music is simply another dialect. Instead, she says that we understand the experience of listening to music in a structuralized manner; this is Raffman's theory of musical ineffability. Her statement that music has its own grammar, which she calls an "M-grammar," is affirmed by music's "possession of (domain-specific) psychological rules."[2] She specifically uses sources from (A) music theory and (B) cognitive science to support this statement.

(A): MUSIC THEORY AND EXPLANATION

A Generative Theory of Tonal Music, a collaboration written by music theorist and composer Fred Lerdahl and linguist Ray Jackendoff, defines the rules and shows the role of M-grammar in our auditory and cognitive perceptions of music. Simply put, a musical grammar does the same thing to musical experience that linguistic grammar does when listening to someone speak — unconscious analysis of a stimulus.[3] We are constantly analyzing what we hear by applying a set of rules. We apply the rules of syntax, pronunciation, conjugation, and others when we listen to spoken word. In music, the rules are different: we hear interactions between rhythm and pitch, or tension and resolution. The Lerdahl-Jackendoff theory states that we "intuitively" apply these rules, which form a grammar, to the experience of listening.

The Lerdahl-Jackendoff theory provides a backdrop for Raffman's explanation of M-grammar. Her explanation follows the basic model of the covering law theory of explanation, as described in Peter Godfrey-Smith's, *Theory and Reality*. According to the covering law theory, to explain something is to "show how to derive it in a logical argument" consisting of both an *explanans* and an *explanandum*.*

Raffman explains M-grammar by stating that musical understanding is the *explanandum* and structural description the *explanans*. Structural description refers to the observable actions that reflect a conscious musical experience. For example, she analyzes the ability of music listeners to report the rhythm, or their changes in verbal and bodily behavior, such as unconscious foot tapping. These behaviors exhibit the human ability to understand music by its components; we recognize the characteristic feelings such as "beat strength, tonal center, harmonic tension, stability, relaxation, and the rest" when we listen to music.[4] Raffman uses the Lerdahl-Jackendoff theory as a sort of general law of music

---

[2] Diana Raffman, *Language, Music, and Mind* (Cambridge, MA: MIT Press, 1993), 41.

[3] Raffman, forthcoming in *Routledge Companion to Philosophy of Music*, 2.

* Put simply, the *explanans* is the thing doing the explaining, and the *explanandum* is the thing being explained.

[4] Raffman, *Language, Music, and Mind*, 49.

theory to explain an unobservable understanding in terms of observables. Her *explanans* consist of testable empirical content, all seemingly true, which bring about the *explanandum*: musical understanding by means of a structure or, more accurately, a grammar.

(B1): COGNITIVE SCIENCE AND EMPIRICISM

The above explanation is a subtle connection between the philosophy of art and the philosophy of science. By explaining her theory with an *explanans* and an *explanandum*, Raffman's argument appears more scientific in character. Moreover, her main reason for explaining M-grammar is to expand upon the relationship between music and language. Her argument in support of this relationship rests on neuroscientific evidence, as detailed in her essay "Music, Philosophy, and Cognitive Science." This contrasts with the Lerdahl-Jackendoff theory, for which the evidence was musical and linguistic "intuition," instead of empirical evidence.[5] In 2006, a series of studies by Akira Miyake and Bob Slevc showed that musical training facilitates the ability to learn a second language. In that same year, fMRI* studies showed that certain musical and linguistic activities trigger the same regions in the brain. Multiple studies from 2006 to 2009 have suggested that the domains of music and language overlap in the brain, since musical and verbal stimuli activate some of the same cortical areas.[6] Such studies further reinforce the connection between music and language that was posited by Raffman.

In her opening chapter, Raffman mentions that her book is meant to be complementary, as opposed to being subversive, to traditional philosophy. Her theory on ineffability is not intended to replace the work of previous music theorists, but instead to contribute to their work. However, her work has an important twist. Where some philosophers claim that science cannot help to describe aesthetic experience, Raffman aims to support the traditional philosophy of musical experience with empirical evidence.[7] She dissects, "from a psychological point of view, what goes on in music perception that could plausibly be described as the acquisition of knowledge that cannot be put into words."[8] Thus, Raffman's theory of ineffability hinges on empiricism.

Her tendency to rely on empiricism, however illuminating, is one-dimensional. She uses scientific evidence to support her philosophical claims, but

---

[5] Raffman, forthcoming in *Routledge Companion to Philosophy of Music*, 3.

* Functional Magnetic Resonance Imaging is a technique that can be used to produce images of brain activation by measuring rates of blood flow across the brain.

[6] Ibid.

[7] Raffman, *Language, Music, and Mind*, 10.

[8] Ibid, 3.

considers that to be the extent of science's role in philosophy. She claims that science can only be useful up until a certain problematic point. For instance, in chapter five of her book, Raffman examines why humans are drawn to ineffable things and why we attempt to verbalize what we experience. [9] She says that we see, feel, hear, smell, and taste many things every day that are ineffable in the same manner that music is, but says that we do not feel the same compulsion to describe some of these things as we do for others. She asks two important questions: why do we direct our attention to the ineffability of music in particular; and, why do we notice what we cannot say? I will refer to these questions as the "problem of ineffability." Since science has not yet given us a solution for this problem, she asserts that philosophy will take over. So, in order to understand these "why" matters, Raffman suggests we need an entirely philosophical explanation. Regarding the role that science can play in explaining ineffability, she says "*science is what you do to a problem until it becomes, for want of a better term, problematical enough to solve by doing philosophy.*" [10]

(B2): COGNITIVE SCIENCE AND INDUCTION

That statement belittles the importance of science by making it the simple precursor to philosophy. It means that science shows "how" something works, and then leaves philosophy to explain "why." Does scientific research serve only to gather data for philosophy to decipher? Perhaps an example will best capture the answer to this question.

In a 2008 psychology study called *Learning, Arts, and the Brain,* Brian Wandell *et al.* tested musical and non-musical children for reading ability (in terms of recitation, comprehension, and retention), finding musically trained children to be stronger readers.[11] Furthermore, his DTI (Diffusion Tensor Imaging)* studies showed that a music education produced a specific change in the development of the corpus callosum, which is a bundle of commissary fibers connecting the right and left hemispheres in the brain.[12] After testing two sets of students, one set having the defining attribute of a music education, the results pointed to one conclusion. Similar to the comparative studies that Raffman uses, Wandell's report arrives at its conclusions by reasoning inductively.

---

[9] Raffman, *Language, Music, and Mind*, 95.

[10] Ibid., 97.

[11] Brian Wandell et al., "Training in the Arts, Reading, and Brain Imaging," in *Learning, Arts, and the Brain*, ed. Carolyn Asbury and Barbara Rich (Washington, DC: Dana, 2008), 55.

* DTI (Diffusion Tensor Imaging) is an MRI (Magnetic Resonance Imaging) technique used by Wandell for three-dimensional visualization of the brain over time.

[12] Ibid., 56.

Assume there is a randomly selected man, whom I will call "Hume," sitting in front of Wandell, and assume that Wandell wants to convince Hume of his conclusions. By showing that one music student is a better reader than his non-musical counterpart, Wandell introduces the possibility of a relationship. Many questions arise as to how legitimate this relationship may be. What if that student was more intelligent than his counterpart? How can we be certain that music was the cause of this relationship instead of some other unseen factor? But one question in particular needs to be asked: what might the relationship between music and reading be? By introducing one trial, Wandell has intrigued Hume with the possibility of a relationship.

After explaining the first trial, Wandell proceeds to introduce more and more trials. These trials no longer arouse the mere curiosity of the first, but instead serve to convince Hume of a relationship between music and reading. Conversely, a plurality of trials makes the possibility of no relationship seem less likely. Starting from scratch, Wandell has introduced many cases where two things occur in succession (a music education and then an enhancement in reading ability). As the cases stack up, Hume is less likely to need a particular case in order to believe the relationship. This results in a theory that a music education contributes to reading ability, and is an example of inductive reasoning.

There is always a chance that Hume will not make this connection. Hume might say that there is no convincing manner to explain a theory just by giving a bunch of cases. Hume could try and find an instance of a "white raven," a case that falsifies the theory, perhaps by finding a music student who performs poorly compared to a non-music student. In spite of this, that student could just be an outlier, such as a student that has a learning disability, or who performed poorly in this case. The majority of science uses and accepts inductive reasoning for the production of justified theories.

After concluding that there is relationship between music training and reading ability, Wandell uses brain imaging to test for neuro-anatomical differences between the musical and non-musical sets of students. He finds, using the same inductive reasoning as before, that there is a relationship between an education in music and an explicit development in the brain (specifically in the corpus callosum). He then reasons, transitively, that 'if X then Y,' and 'if Y then Z,' X must be related to Z. If there is a relationship between a development in the brain and music training, and a relationship between music training and reading ability, then the development of the corpus callosum may be a key connection between music and language. Wandell's expansion is not unusual in science, since many scientists attempt to expand upon their findings in an attempt to explain more and more.

Recalling the question I asked earlier, regarding whether or not science is

just data to be interpreted by philosophy, the answer is a resounding *no*. This can best be shown by examples from science: with induction, scientists make what I call an "inductive jump" from their results to their conclusions. Brian Wandell not only jumps from a set of trials to a tentative theory, but he expands his experiment to provide further explanation with induction.

When does science end and philosophy begin? When scientists expand upon their findings do they become philosophers? Is an inductive jump a movement from science to philosophy? If so, then Raffman's claim may indeed apply to all situations. This is a categorical distinction that Raffman does not make in her essay, which leads me to my point. Wandell's peers still call him a scientist when he uses induction, and so long as induction is considered justified in science Wandell is making a logical scientific claim. Unless Raffman clearly makes a distinction between philosophy and science, I will continue to interpret her claim assuming that a scientist remains a scientist during an inductive jump.

Under this assumption there is no reason as to why a scientist might not be able to explain the problem of ineffability, and there very well could be a forthcoming psychological explanation. An individual's attention to things that they cannot explain is a behavior-related problem. With the rapid progression of neuroscience and cognitive science, I would not be surprised if humans eventually understood the reasoning behind attention and behavior toward abstract concepts such as ineffability. Over time, scientific developments allow scientists to explain increasingly more difficult problems.

INEFFABILITY & INCOMPATIBILITY

Through use of inductive jumps, scientists have the potential to explain their findings. Throughout history, science has explained things that people deemed impossible; for instance, consider Albert Einstein's theory of general relativity. Einstein's theory successfully predicted the Hyades star cluster to be visible during the solar eclipse of 1919. Although the stars were actually located behind the sun, they appeared to be next to the sun during the eclipse. Einstein explained that since light projected from the Hyades star cluster was curved by the gravity of the sun (an axiom of relativity), the starlight that should have passed by the earth was instead projected onto our planet.[13] I introduced general relativity as an example of science making progress. If there are limits on science, this progress must then exhaust itself at some point. In Raffman's claim, this is the point at which philosophy takes over. Her claim implies an *extremely* pessimistic view of scientific realism.

---

[13] Richard Ellis et al., "1919 Eclipse Revisited," ed. Sue Bowler, *Astronomy & Geophysics* 50 (2009): Apr. 12-15, 4.

Godfrey-Smith explains that a scientific realist has a specific attitude towards science's representation of the world. He says there are two different attitudes within the bounds of scientific realism – optimism and pessimism – and he describes them as follows: an optimistic scientific realist is confident that science will figure out how the world works and will ultimately find the correct theories for everything, while the pessimist thinks that, though we may attain the right theories, we are overconfident in our abilities to understand and we get confident too quickly.[14] Note that the pessimist thinks it is science that cannot immediately get the right theories, but this does not apply to other fields of research. Raffman is a pessimist by this definition, since she thinks scientific research cannot find a solution after a point, and that philosophy is the alternative solution. By placing limits on science, though, Raffman becomes a special kind of pessimist.

Godfrey-Smith mentions that there is a special pessimist that thinks it is impossible for science to get the right theories; he calls this *extreme* pessimism.[15] Up until a problematic point, Raffman's view of science does not contradict the definition of scientific realism. Since she also talks about the cognitive science of music promising countless possibilities, it would appear that Raffman is a scientific realist. After a problematic point, however, she implies that science cannot get the right theories to solve the problem of ineffability, and she transforms into an *extremely* pessimistic scientific realist. Of course, some questions fall outside of the scope of science. But Raffman has based her theory of ineffability on science, so why can't science extend to explain a problem of ineffability?

Godfrey-Smith states that extreme pessimism is incompatible with scientific realism. One of the central beliefs of a realist is that science can get theories right, on some level. Therefore, an anti-realist believes it impossible for science to generate correct theories about the world. Raffman's claim actually lies in the grey area between realism and anti-realism. This sort of fluctuation between views makes for a weak scientific argument.

SILENT BITE

Raffman contradicts her claim on ineffability in an earlier chapter of *Language, Music, and Mind*. After the birth of the cognitive science of music in the 1980s, there was a great deal of new empirical research being done and many discoveries in psychology.[16] Raffman recognizes this in the following

---

[14] Peter Godfrey-Smith, *Theory and Reality* (Chicago: University of Chicago Press, 2003), 177.
[15] Ibid.
[16] Raffman, forthcoming in *Routledge Companion to Philosophy of Music*, 1.

passage and also stresses how these discoveries created many opportunities for philosophy, but philosophers failed to capitalize on them.

> Certainly there are those who have recognized the need for psychological theory in the philosophy of music, but to my mind they have not followed through. That our forebears should disappoint in this respect is hardly surprising, for only recently has a *bona fide* psychology of music been available. Music theorists have long supposed that musical experience and behavior are systematically related to structural regularities in heard pieces, but only with the advent of so-called cognitive theorizing have the psychologists appropriated music perception as an object of scientific study.[17]

This says that philosophers failed to capitalize on a *bona fide* psychology of music, which means philosophy was silent before the *bona fide* psychology was made available. By *bona fide*, she also means that science of this type was not done honestly for a period of time. So, science was having problems (with legitimacy) and, because of those problems, philosophy was silent. This period suggests the opposite of Raffman's claim that I have been discussing, giving rise to a contradiction between chapters one and five of her book. The above passage from chapter one shows that philosophy was caught at a standstill and science found its own solution. The silence on the part of philosophy comes back to bite Raffman, since in chapter five she suggests that philosophy is the solution after a problematic point.

IN CONCLUSION:

To reiterate Raffman's claim: science can solve a problem until that problem becomes problematic enough to solve by doing philosophy. I do not think that this statement is always wrong. I think there is some merit to this claim, as I am sure it has been the case at some point in history. In most cases, however, it does not necessarily apply. In the case of musical ineffability, I do not think it applies.

Her claim insists that philosophy and science are ordered. Why must one follow the other? She also never provides a distinction between philosophy and science. At which point in an explanation does science stop and philosophy take over? Instead of getting caught up in problems associated with succession,

---

[17] Raffman, forthcoming in *Routledge Companion to Philosophy of Music*, 3.

I suggest an alternative relationship between philosophy and science. Although the two do not exist independently of one another – indeed, each has been shown in this paper alone to have an impact on the other – both philosophy and science progress on distinct paths. For the majority, they are two separate fields with two separate groups of researchers, but they certainly influence each other and there is some overlap. This reminds me of the process of neurulation: during embryonic development, there is a period when animal embryos are composed of several different layers.* Each layer progresses towards its own "goal" (in forming the nervous system), but constantly sends and receives inductive signals, which affect those "goals," to and from the other layers. I think this is closer to the truth than Raffman's view, but the real relationship is far more complicated.

The boldness of Raffman's claim comes from its oversimplification. I could find objections to it because it is a very narrow depiction of a very broad topic — the relationship between philosophy and science. That said, I did not pursue her claim with such determination in order to prove a point in the philosophy of music. Rather, I meant to show the potential utilization of the philosophy of science for assessing a non-scientific philosophy. For instance, the theories of great philosophers are the basis for modern arguments: the covering law theory of explanation appears to be the foundation for Raffman's argument for M-grammar. Also, the application of inductive reasoning in theories of both philosophers and scientists shows that many years of philosophical debate did not go to waste. Raffman's argument for a relationship between music and language is steady, in great part due to the strength of induction, both in her own and in others' arguments. Arguments can be weakened or strengthened based on their ability to adhere to certain philosophical doctrines. This is clear in Raffman's claim and its inability to fit into the doctrine of scientific realism: the claim contradicts itself. Thus, groundwork and terminology laid down by philosophers clearly produce a more universal understanding of an argument.

The cognitive relationship between music and language, if real, holds many promises for future developments and understanding. Additionally, the ease by which doctrines from the philosophy of science fit over theories in the philosophy of music shows a clear convergence between the two disciplines. Perhaps there is a greater connection between language, music, and science than yet realized.

---

* For curiosity's sake: the process of neurulation results in the formation of the ectoderm, endoderm, and mesoderm, which develop into different parts of the animal body, after receiving inductive chemical signals from the notochord.

BIBLIOGRAPHY

Ellis, Richard et al. "1919 Eclipse Revisited." Ed. Sue Bowler. *Astronomy & Geophysics* 50 (2009): Apr. 12-15. Web. 20 Nov. 2010.

Godfrey-Smith, Peter. *Theory and Reality*. Chicago: University of Chicago, 2003.

Raffman, Diana. *Language, Music, and Mind*. Cambridge, MA: MIT Press, 1993.

Raffman, Diana. "Music, Philosophy, and Cognitive Science." Forthcoming in *Routledge Companion to Philosophy of Music*, 2011. University of Toronto, Web. 7 Nov. 2010.

Wandell, Brian et al. "Training in the Arts, Reading, and Brain Imaging." *Learning, Arts, and the Brain*, Ed. Carolyn Asbury and Barbara Rich, 51-59. Washington, DC: Dana, 2008. Web. 11 Nov. 2010.

FURTHER READINGS

Forgeard, Marie, Ellen Winner, Andrea Norton, and Gottfried Schlaug. "Practicing a Musical Instrument in Childhood Is Associated with Enhanced Verbal Ability and Nonverbal Reasoning." Ed. Tecumseh Fitch. *PLoS ONE* 3.10 (2008): 1-7.

Spelke, Elizabeth. "Effects of Music Instruction on Developing Cognitive Systems at the Foundations of Mathematics and Science." *Learning, Arts, and the Brain*, Ed. Carolyn Asbury and Barbara Rich, 17-49. Washington, DC: Dana, 2008.

Pearl, Jonathan G.S. "Song, Speech, and Brain." University of California, Santa Barbara. *Music & Language Studies*. 3 Sept. 2001. Web. 9 Nov. 2010. <http://www.musiclanguage.net/>.

Eliminating Causal Theories of Mental Content:

Detection, Content-Fixing, and Reference

Peter Tan
University of Illinois, Urbana-Champaign

1. INTRODUCTION

The *language of thought* (LOT) hypothesis asserts that the scheme of human mental representation is a language-like system wherein complex mental representations are syntactic arrangements of semantically primitive mental representations. A *causal theory of mental content* (CT) is any sort of theory of mental content that says the semantic content of primitive mental representations is causally fixed to some property or instance of it. I will present and defend Robert Cummins' argument that CT entails LOT. The thesis of this paper, however, is to show that CT fails to meet its own expectations for what the content of semantically primitive mental terms is like, and that ultimately, CT is a self-defeating theory of mental content.

The argument I present runs as follows:

(1) CT entails LOT;
(2) CT cannot explain that semantically primitive mental terms refer to their causes;
(3) CT is insufficient as a theory of mental content.

This paper will proceed in two stages: first, I will present Robert Cummins' 1997 argument in support of (1) and briefly defend objections to it; then, I will defend (2). The critical caveat to note on (2) is that LOT requires that semantically primitive terms refer to their causes – or so I will argue. My argument is hence a modus tollens on CT and LOT. It should be noted, however, that my intention in this essay is not to undercut LOT directly. There may very well be a theoretical framework for fixing the content of mental terms empirically that does not entail LOT; there may very well be a way of holding LOT that does not require CT. I only aim to argue that CT and LOT together in their present state have serious shortcomings in appropriately explaining mental content and – by extension – mental representation as a whole.

2. CT ENTAILS LOT

CT seeks to provide an account for how the contents of semantically primitive mental representations are fixed. Central to all causal theories is the idea that causal connections between external properties and mental representations determine the semantic content of primitive mental terms. The many proponents of CT have each defined the proper content-fixing relationship in their own way. In its simplest form, however, CT supposes that "the content of a primitive r in a system $\sum$ is the property P if there is the right kind of causal connection between

instantiations of P and tokening of r by Σ's detectors."[1] A mental token $|x|$ means x because an x caused $|x|$. Aside from each primitive term having the right kind of content-fixing relation to its members, though, the coinage of semantically primitive terms is supposed to be arbitrary. This is true in the following sense: the primitive term that is coined when a certain property P is detected need not bear any resemblance to its cause, or any direct indication of what its content is. If P can be detected at all, LOT can churn out any arbitrary symbol when it detects P. So, any primitive term can represent any detectable property.[2]

CT seems plausible enough, *prima facie*, as a theory of mental content. Since it asserts that all mental content is causally determined, however, CT is a slippery slope to the idea that even the most complex of mental representations are atomistic, and can be constructed and manipulated syntactically.* If mental content is causally fixed, as alleged in CT, the content of all mental terms must either be detectable, or otherwise be available through combining some terms that do have detectable content. A |cat on a porch| might very reasonably be caused by a cat on a porch, but it is hard to believe that a |cow jumping over the moon| could be caused by a cow jumping over the moon. It is its closeness to LOT that makes CT a tough bullet to bite. There may in fact be a way of explaining the intentionality of mental content that does not entail LOT; but as far as CT goes, I argue that in order to accept CT as a true theory of mental content, LOT must also be true.

Robert Cummins' argument that CT entails LOT runs as follows:

(C1) CT entails that there are finitely many semantically primitive representations in the scheme of mental representation.

(C2) Since mental representation needs to be productive, there must be a scheme of combining semantically primitive representations into complex representations that obtain their meanings as a function of syntax and their constituents' content.

(C3) The type of scheme described in (C2) is LOT.

He invites us to consider that CT requires a detector for the content of each primitive representation. For every property that the LOT wants a primitive term for, there must be a mechanism that detects instances of it. He asks: can a finite system – such as the human brain – that can hold only a finite number

[1] Robert Cummins, "The LOT of the Causal Theory of Mental Content," *Journal of Philosophy* 94 (1997): 535.
[2] Ibid., 539.
* Cf. Jesse J. Prinz, *Furnishing the Mind: Concepts and Their Perceptual Basis* (Cambridge, MA: MIT Press, 2006) pp. 89-100.

of detectors, detect infinitely many target properties and maintain semantically primitive terms for each of them?

2.1 THE FINITENESS OF CT TERMS

Cummins' reply is "no." He hypothesizes a mechanism – an array of photosensors – that detects squares, in any size, orientation, or location on the array. Now, this indeed is a finite representational system, since the plane of photosensors is not infinitely large. Let's suppose (as the causal theorist would want to) that any square pattern on the array is a primitive representation of the property of squareness. But that's not right – as Cummins points out, CT would say that each figure represented on the array only represents an individual square, not the property of squareness proper. In order for CT to actually detect the property of squareness in each figure, instead of just seeing a multitude of four-sided figures, it has to abstract from each primitive representation. It has to know squares: it has to have a theory of what a square is. The same applies for any other shape or property that CT tokens primitives for, since CT requires the detection of *P* before a primitive term containing *P* can be tokened. Since there is only a limited amount of memory in a system like the one described above, and in the human brain, the number of primitives that represent target properties will be finite. Cummins writes:

> CT allows for a scheme in which the set of primitives is unbounded in the sense that, given enough memory, another can always be added. But it does not allow for a scheme in which the number of primitives is infinite.[3]

So, CT entails that there are finitely many semantically primitive representations.

Though the range of primitive terms coined by CT is necessarily limited in number, however, human mental processes ought, in principle, to be capable of generating an unbounded number of semantically distinct complex representations. This is what I take 'productive' to mean. And since the building blocks for semantically complex representations – primitive terms – are limited in number, there needs to be some kind of combinatorial mechanism through which the semantic content of the complex can be achieved through manipulating the primitive. In other words, CT requires a representational system such that every complex representation is a combination of some primitive representations, and that the semantic content of any complex representation, is a function of the content of its constituent terms, and the way in which they are arranged. The sort of representational scheme CT requires, Cummins concludes, "is LOT as near as makes no difference."[4]

---

[3] Cummins, "The LOT of the Causal Theory of Mental Content," 540.
[4] Ibid., 541.

2.2 *P*-DETECTION AND THEORY-MEDIATION

A committed CTer may try to assert against (C1) that the human representational system can have infinitely many semantically primitive terms. The first premise of Cummins' argument hinges on the assertion that in order for my representational system to detect a property *P*, it needs to have a theory of *P*. I need to have a theory of squares, of cats, and so forth. Cummins calls this elsewhere the *nontransducibility of distal properties* (NTDP). Since my detectors need an explicit theory of what a property *P* is in order for *P* to be properly detected, and CT needs proper detection of *P* in order for a semantic primitive of it to be tokened, the number of semantically primitive representations will be limited by the number of theories of various properties *P* that my representational scheme can hold. But if primitive terms are some kind of abstracta and not some quasi-mechanical output, as Cummins seems to suggest, there is no need for the whole system to operate in a LOT-like manner to be 'productive' in the manner described above. Primitive representations are finite in number because tokening them requires building theory-mediated detectors in a finite system. I suppose, then, that the CTer who makes this objection will want to say that the tokening of primitive terms is not originally theory-mediated. The CTer must find a way, however, to maintain that semantically primitive terms can be coined reliably – that is to say, without mental misrepresentation. So, the question becomes: how do I build a reliable *P*-detector?

Let's suppose I want to build a fox detector – not a fox hair detector, nor a fox whisker detector, but a whole fox detector. Now, suppose I find myself in the woods and I encounter an animal – I detect that it is reddish-orange, about the size of a domestic dog, and quick on its feet. Have I successfully detected a fox? Maybe. I will have successfully detected a creature with many physical characteristics of a fox, though it is equally likely it will have been just a reddish dog. In order to *reliably* detect foxes, however, I need a detector knowing that when those proximal stimuli – color, size, and quickness – occur in concert, they indicate a certain distal stimulus, namely, a fox. If indeed I want a fox detector and not a fox-and-dog detector, I need a detector that has a theory of foxes.* The proper detection of a certain distal stimulus *P* is, in fact, a function of my proximal stimuli correctly satisfying what my detectors say a *P* is. In order for the CTer to believe

---

* The sort of problem described in this paragraph is what Jerry Fodor has called the 'disjunction problem'. Here's what it is: CT says that a mental token represents its cause. So, if the token |fox| is caused not by a fox, but by a dog, we are forced to say that |fox| here represents the property of being a dog. It becomes the case that both foxes and dogs can cause |fox|s, and we are forced to say that the token |fox| represents the disjunctive property of being a fox-or-dog – but |fox| is supposed to result when and only when, and because, a fox caused it. It is because of the disjunction problem that Fodor himself accepts NTDP. His proposed solution to the problem is something called 'asymmetric dependency'. This idea proposes that I can mistakenly detect dogs IFF I can reliably detect foxes. But the fact that my detectors can reliably detect foxes (and the fact that my mistakenly detecting dogs depends on my reliably detecting foxes) requires that my detectors do actually know what a fox is. So, NTDP (Cf. Fodor, 1990).

that content-fixing causal links can be reliable, then, he must believe that all of his many *P*-detectors can reliably interpret the content of proximal stimuli. So, the initial tokening and content-fixing of semantically primitive mental terms must be theory-mediated, and the number of semantically primitive terms a CT can churn out *is* finite. Therefore, (1) CT entails LOT.

3. ON LOT TERMS AND REFERENCE

Primitive terms in LOT are supposed to refer to their causes in the same way that words in public language have their referents. When I watch *Casablanca* and have the thought that |Humphrey Bogart kissed Ingrid Bergman|, the mental terms |Humphrey Bogart| and |Ingrid Bergman| refer to their cause – namely, them on the screen and their individual persons – in the same way that an English utterance of "Bogart" or "Bergman" intuitively refers to the same. It is precisely for this reason that CT may be so plausible *prima facie*. It is appealing to say that my mental terms have content that actually is in the external world, caused by the external world, and acquired independently from other mental content of mine.* And it is also for this reason that CT's shortcomings as a theory of mental content become evident.

CT requires that the content of any primitive mental term $|x|$ directly refers to its external causes. But CT also requires NTDP for the content of any primitive mental term $|x|$ to be properly fixed – at least in its initial tokening. CT-cum-NTDP therefore allows $|x|$ to be tokened if and only if I have an x-theory and something y satisfies it in the appropriate way (bearing in mind that those things y will not always be the actual xs that I want). Additionally, recall that any x-theory I have will by definition be composed of whatever I happen to know about the relevant x. The content-fixing of any primitive mental term, then, in fact relies more on my explicit theory on the detection of the content of that term, than on causal links from having detected some actual x. But then, $|x|$s aren't necessarily caused by xs. They are, however, necessarily caused by my NTDP-informed internal states. Ultimately, then, this means that NTDP is the real source for the tokening of primitive mental terms – at least more reliably than CT's alleged causal links. So a mental term will not necessarily represent its external causes, but it will always represent its internal causes. CT wants its terms to have external content and causal origin, but fails to provide a consistent account of how that can be achieved. I say that in the end, CT cannot explain that mental primitives refer to their causes.

* Portions of the following sections revolve around the wide/narrow mental content debate. I simply assume that externalism about mental content is true. For a more up-to-date defense of externalism, following Putnam's classic 'Twin Earth' defense, I refer the reader to Paul Noordhof (2006), Jason Bridges (2006), and Falvey and Owens (1994), among many others.

In the rest of this portion of the essay, I defend what has just been said.

It may seem from the outline of my argument above that my elimination of CT centers on a certain way of handling cases of mental misrepresentation. This is in fact not the case. I argue that CT's own defense against mental misrepresentation – which heavily involves NTDP – in fact weakens the CTer's case that the content of primitive mental terms can be causally determined. My argument against CT is that it fails to meet the requirements of its own theoretical framework. This framework, of course, is LOT. Of course, whether LOT exists is a contentious issue – all Cummins' argument shows is that some minimal form of LOT is entailed by CT. I make no judgment on whether LOT actually exists; I argue, however, that the same reasons that cause CT to entail LOT dismantle CT's coherence with LOT and CT's own attractiveness as a theory of mental content. It is necessary to show that CT entails a LOT framework for human mental representation, because CT's demise comes from its incompatibility with LOT's own requirements on mental content. CT is self-defeating.

There is no need here for me to support that CT requires NTDP, or argue for NTDP's constraints on the detection of proximal stimuli; both of those have already been accomplished. I imagine, however, that what needs defense now is my assertion that CT relies on NTDP (which is to say, internal causal links), rather than external causal links, to fix the content of semantically primitive mental terms. After all, CT is only a satisfying account of mental content if the causal content-forming links are indeed what it says they are. And after all, the *only* good reason CT has to say that *any* $|x|$ refers to some x is because an x ought to have literally caused $|x|$ to appear. If it can be shown that those causal links are in fact rather weak, or that they do not actually exist, then CT can be weakened, and even eliminated. I will show exactly that.

3.1 CT AND REFERRERS

CT wants a scheme of mental representation in which semantically primitive terms are direct referrers to external properties. The reason for this is an observation I previously made: namely, that LOT – and indeed also CT – is appealing much because it allows for a scheme of mental representation in which mental content is external. The difference between reference and other sorts of supposed content-fixing relationships (at least those concerned with intentionality) is a matter of how cognitively expensive each is. I am willing to draw the line here: reference is cognitively cheap (that is to say, epistemically). I am not required to know anything in particular about something in order to refer to it. I can say, "I cannot tell the difference between beeches and elms," without actually knowing what a beech is, or any characteristics beeches typically

have. I can successfully refer to beeches without knowing anything other than the fact that beeches exist.* On the other hand, I can know the whole world of information there is to know about beeches, and my utterance 'beech' (as well as my mental term |beech|, according to the CTer) will be no more successful in denoting a particular genus of deciduous tree, than if I knew nothing.** I would argue that a mental referrer is any semantically primitive term whose content ought to be possible to be fixed wholly ostensively. I should be able to point at a beech, refer to it, and use it in LOT formulations – "this thing here is rather tall" – without any sort of sophisticated beech-theory, as it were. Indeed LOT, though a private language, wants terms that refer to public objects.

But the causal links that (CT says) link my mental primitives to their causes are far weaker than the CTer would like to presume. Take the familiar case of Little Albert: when presented with a white rat, he was conditioned to have a phobic reaction. The causal theorist must admit that upon being first presented with the white rat, Albert had to construct a white-rat-theory – since, after all, the detection of distal stimuli is theory-mediated. This is easy enough to accept. As it happens, however, Little Albert's white-rat-theory came to include a terrifyingly loud noise, in addition to more typical white-rat-theory contents of whiteness, timidity, and furriness. After constructing his white-rat-theory, Albert would react to the white rat phobically, even if no noise was presented. Now, the CTer ought to say that no such reaction should occur: after all, Albert's |white rat| refers to the actual white rat on the table, in whose presentation there is no loud noise. The fact that he *does* react phobically presents serious problems – it shows that Albert's |white rat| in fact refers to and contains his noisy white-rat-theory, not the noiseless white rat on the table. The initial rat-and-noise combination may indeed have caused his rat-theory to be formulated in a certain way, but his conditioned responses show that NTDP – his internal white-rat-theory – was the actual cause of his having fixed the content for that term, and so also that term's actual referent. So, CT's causal links fail to fix the content of primitive mental terms as it says they should.

---

* Is this also necessary? Perhaps not. Or, more accurately, all that is necessary is having my circumstances be such that 'beech' is not a vacuous term. There is a certain natural kind, namely a beech, which the English-speaking dendrological community has been kind enough to name. I can utter 'bxxoqpwe' and claim that it refers to a beech, but that will never fly. Likewise I can utter 'onomatopoeia' and claim that it is a vacuous term (or that it refers to beeches), but my circumstances dictate that I will be wrong on both counts. (Cf. Tyler Burge, 1986; Speaks, 2006.)

** One objection to my argument, so obvious I must address it here, is the suggestion that we can, we ought to, and we often do in fact know much about the content of our mental terms. And of course I agree with this – it'd be strange not to. My only submission is that we don't *need* to know about our mental content for our terms to be meaningful as we think they ought to be: rather, linguistic meaning needs to be prior to mental content. As Jeff Speaks relates, the problem with CT is that it "is an attempt to give an account of the content of a mental representation in terms of its occurrence in a thought-state; but because there is no guarantee that if a property occurs in a proposition, then the truth of the proposition entails that the property is instantiated; there is no guarantee that, even if we restrict ourselves to true thoughts, it follows that there is a reliable correlation between the presence of a mental representation in a thought-state and the instantiation of any property at all" (Cf. Speaks, 2006, p.437).

Likewise, the causal links that (CT says) link my mental primitives to their causes often do not even exist. It becomes clear upon close inspection, for example, that my tokening |fox| after seeing a red fox does not actually refer to the animal I just saw. It only indicates that whatever I just detected was reddish, dog-sized, and light-footed. Recall that I token |fox| just in case my fox-theory has presently been satisfied – but I will still token |fox| in cases where no fox is actually present. In other words, I token |fox| because |fox| does not actually contain a fox; it contains my theory of foxes. This is the real reason why I can mistakenly token |fox| after seeing a red dog: my term |fox| does not refer to any external fox; it only indicates that I've detected *what I think a fox to be*. And since I need a theory of what a property x is before I can fix any term to contain that property, any $|x|$ I can token will at best indicate that I have an operable x-theory. The fact that I might token an $|x|$ after seeing a y is extremely problematic for the CTer: the content of my $|x|$ will be fixed as usual, and $|x|$ reliably tokened, despite the complete absence of a real x. NTDP tokens mental terms reliably even in the absence of the appropriate external stimuli, and CT cannot. So NTDP is the real cause of the tokening of primitive mental terms. So, CT's alleged causal links between xs and my $|x|$s don't actually exist.

So CT does not get direct referrers. Recall that according to CT, in order for me to token a semantically primitive term, I need to be able to detect the content of that term. And, to be able to detect it, I need to build a detector expressly for that purpose. But, if I am to build a detector in order to reliably detect a certain property x, I need a theory of what that property is and what sort of proximal stimuli I can expect to accompany it. CT says that in order to fix the content of a LOT term, I should know about it. But CT also requires that I cannot fix a term's content at all unless I do know about it! CT-cum-NTDP ends up being a framework in which the supposed content-fixing external causal links cannot actually exist. So the terms that CT tokens cannot actually represent their external causes. CT cannot token directly referring primitive mental terms; so, (2) CT cannot explain that semantically primitive mental terms refer to their causes. And, since CT and LOT require referrers, (3) CT is insufficient as a theory of mental content.

4. SOME OBJECTIONS CONSIDERED

4.1. OBJECTION 1

Robert Cummins' own argument against CT is that CT fails because the requirement that the content-determination of semantically primitive mental

terms is explicitly theory-mediated entails a circularity in CT's representational framework. Taking in mind that the stock of primitive mental terms innate to the human mind is very limited, Cummins points out that having to formulate a theory that "(A) Cats have whiskers (B) Cats have four legs (C) Cats have fur" requires having a term, |cat|, to begin with. But CT says I cannot have a |cat| until I have the ability to detect cats through an explicit cat-theory.[5] CT is circular.

Robert Rupert (2001) attempts to rebut Cummins' argument directly. His argument centers on a revised version of the strict notion of innateness found in Cummins' text, drawing heavily on advances in developmental neurobiology suggesting that "innateness" might be more appropriately understood as "early emergence."* While Cummins is right to argue, he says, that the scarcity of mental terms present at birth shifts the weight of the content-fixing ordeal towards NTDP, it is not true that the lack of a |cat| (to use the previous example) in my stock of innate terms, precludes my ability to detect cats. He summarizes his conception of mental innateness, and its consequences for Cummins' circularity argument, as follows:

> Early human development comprises an ongoing process of neural generation, growth, and death and synaptic generation, strengthening, and weakening which results in the emergence over time of functionally coherent cell assemblies. If this picture is correct, then at least some of the time after birth, the human brain/cognitive system generates, or 'coins', new LOT terms, individuated nonsemantically; it generates new circuits or cell assemblies of the same sort as those we describe when we characterize LOT terms in the language of the neurosciences. It is possible, then, that the subject's early theory of catness include a newly coined LOT term lacking the content cat, perhaps lacking content altogether.[6]

It seems highly unintuitive that there might be LOT terms, innate or otherwise, that totally lack semantic content. But this view on innateness includes that the range of native primitive LOT terms might include nonsemantic ones that

---

[5] Cummins, "The LOT of the Causal Theory of Mental Content," 537.

* In understanding this revision, it is important to understand that Rupert also draws here a distinction between innateness and unlearnedness. He suggests that while Cummins supposes that our innate LOT terms ought to have their contents fixed in a way not requiring NTDP, those kinds of content-fixed native terms ought to be called 'unlearned', since their content is unlearned, while a more appropriate picture of 'innateness' is that what we think of as 'innate' terms, are occurrences arising in the neural substrate after certain kinds of infantile and early childhood interaction with our environment. It is these innate, emergent terms that Rupert countenances that serve as the nonsemantically individuated LOT vehicles in the initial formulation of explicit theory-mediated detection, expounded upon in this section (Cf. Rupert, 2006, pp. 516-520).

[6] Robert Rupert, "Coining Terms in the Language of Thought: Innateness, Emergence, and the Lot of Cummins's Argument Against the Causal Theory of Mental Content," *Journal of Philosophy* 98(2001): 522.

are differentiable from other unlearned terms and learned non-native terms.* It is this possibility for the nonsemantic individuation of terms which, Rupert says, undermines Cummins' argument: nonsemantically individuated LOT terms can serve as vehicles for theory-mediation to begin. He writes,

> The initial theoretical axioms that include the nonsemantically individuated LOT term we call 'cat' [should] more properly be thought of in the following way: (A) *t* (a newly coined, nonsemantically individuated LOT term) have whiskers; (B) *t* have four legs; (C) *t* have fur. Cummins's fundamental objection to CT is thus dissolved. Take learned concept *C*. Given CT and NTDP, a subject cannot possess a concept with the content *C* unless she possesses a nonsemantically individuated vehicle '*C*'; the vehicle must be present to hold a place in the subject's theory of *C*s. But '*C* ', characterized nonsemantically, does not have to have content *C* when it first appears in the axioms constituting the theory that mediates the fixing of content *C* to vehicle '*C*'. The alleged circularity of CT vanishes.[7]

If Rupert's rebuttal of Cummins' argument is in fact sound, it poses a serious threat to my argument in that NTDP would play a far less significant role in fixing the content of primitive LOT terms than I (or Cummins) have argued. The crucial premise in my argument is that explicit theory-mediation, rather than external causation, is the source of the content-fixing of primitive LOT terms. But if the content of semantically primitive LOT terms does not rely on explicit theory-mediation to be fixed to certain properties, but instead becomes fixed gradually and naturally as a result of the emergence of new neural substrates, my argument fails because, contrary to what I have argued, the content-fixation of my $|x|$s will actually have been caused by *x*s – or, at least, many *x*s throughout the duration of my theory-building. And so while CT would still entail LOT, it would meet LOT's expectations regarding semantic externalism, and my argument would fail.

---

* There is more to be said about this: Rupert's view is interesting in particular for noting that primitive LOT terms should be able to be differentiated without appealing to their semantic content. This is to say, there are ways of nonsemantically individuating primitive LOT terms. This might be achievable through some syntactic function, bearing in mind that mental reductionism of the LOT variety often contends that there are LOT analogues of grammatical categories such as nouns, verbs, and adjectives. In formulating his argument, Rupert supports only that there are nonsemantic individuation criteria (though not necessarily syntactic). He writes, "when the LOT theorist says, 'That item has content *C*', I expect the LOT theorist to be able to describe what that item is, in some way other than by reference to *C*. Note that this condition does not demand anything nearly as strong as might be demanded by a species-wide type-type identity criterion, which is often construed to require that a given LOT term individuated nonsemantically should be found in all the various members of the same species and found to have the same content in all members of the species" (Cf. Rupert, p.516, footnote 30).

[7] Rupert, "Coining Terms in the Language of Thought," 523-24.

I am actually willing to accept Rupert's revisions on the definition of innateness, and also accept, to an extent, that the content-fixing of primitive LOT terms might be neurologically mediated. But, I think that his claims that there are nonsemantic LOT terms fail – I do not think that there can be LOT terms that genuinely lack semantic content. Let's suppose there actually are LOT terms that lack content whatsoever. Let's also suppose, as he does, that there is a basis in developmental neurobiology for the emergence of nonsemantic LOT 'vehicles' that serve to help construct explicit NTDP theories.[8] The obvious question, then, is what the nonsemantic individuation criterion for those LOT terms actually is. I show here that Rupert's reply undercuts his argument.

Since he supposes that these nonsemantically individuated LOT terms arise in the brain in infancy or early childhood, he accepts the claim that:

> Environmental instruction causes the growth of the very neural resources that are needed to represent aspects of problems the subject will solve using those resources: the specific character of experience catalyzes the growth of the neural resources needed to process more effectively the input that triggered the development of those resources.[9]

But then, doesn't the emergence of a particular kind of neural resource over the emergence of another actually, implicitly have semantic content? In other words, wouldn't using the environmental emergence of a nonsemantically individuated LOT term to individuate it suggest that there is a particular reason it is *that* particular nonsemantic term, instead of another one that could just as likely have emerged in the same environment? So there are no genuinely nonsemantic individuation criteria: every way that we might individuate a nonsemantically individuated LOT term relies on what the environment of its emergence was, and reveals factors in its neural formulation that will, for better or worse, play a role in its eventual content-fixing.

So there are no genuinely nonsemantically differentiable primitive LOT terms. Rupert's argument fails: he proposes that some term, *C*, characterized and individuated nonsemantically, might be able to serve as a vehicle for the building of an explicit *C*-theory. But the existence of a *C* at all actually shows that there is some implicit semantic content already in that LOT term. The emergence of a nonsemantically individuated, non-theory-mediated term still requires some amount of theory-mediation. Rupert's claim that CT is not circular ends up being circular itself.

---

[8] Rupert, "Coining Terms in the Language of Thought," 517.
[9] Ibid, 519.

4.2. OBJECTION 2

Another reader might object that my argument can be dodged by supposing that primitive LOT terms might not actually have to directly refer to the world, in the linguistic sense. This reader says, CT is and always has been in the business of representing things, not referring to them. In defense of this sort of view, Marius Usher (2001) offers an account of mental misrepresentation, allowing causally-created mental terms that do not represent the object or property that actually caused them, but only what is *most likely* to have caused them. This raises a problem for my argument, since my defense of (2) centers on the intuition that the only reason CT has to say that a primitive mental representation refers to its cause is because of actual causal links, not probabilistic ones. If, from the outset, CT does not need direct causation to fix the content of its primitive terms, CT fails to be self-defeating in the way that I've argued. This kind of theory would alleviate the worry about external reference not being satisfied, because it tries to rewrite what mental representation (and misrepresentation) actually is.* Usher puts it this way:

> The role of mental representations may...be to provide, not faithful access to the class of objects causing an act of perception, but rather a statistical inference (or hypothesis) of what type of object *could be* causally involved...In other words, when a concept is tokened, what is represented is not the type that caused the mental state but the type that is the *most likely* to have caused it. This is consistent with Dretske's original idea that mental symbols represent what they carry information about.[10]

A statistical account of causal content such as this has, of course, revisions on what mental content looks like: following Ruth Millikan, Usher adapts a sort of non-referential account of mental content. He writes,

> Millikan (1998) argues that concepts are to be individuated by the capacity to identify exemplars rather than by a

---

* A quick note: I recognize that some readers might take issue with my earlier presumption that externalism about mental content is true. However, though I do take externalism to be true, no premise of mine requires that. I do not consider the possibility that externalism regarding mental content might be false to be an objection to my argument. After all, I do not argue that CT fails because externalism is true; I've only argued that CT fails because – together with LOT – *they* require externalism to be true.

[10] Marius Usher, "A Statistical Referential Theory of Content: Using Information Theory to Account for Misrepresentation," *Mind and Language* 16 (2002): 316.

> description of their properties (i.e., the concept 'cat' is individuated by the ability to tell cats from non-cats, rather than by being able to list properties such as fur, meowing, etc.). These properties are, according to Millikan, secondary to the referential character of concepts. They are acquired later during human development (infants acquire linguistic representation of substances, such as 'animal' and 'food' much earlier than they acquire linguistic terms for their properties), and the ability to identify a substance is needed in order to confer to it a set of properties. This approach is also consistent with Fodor's (1998) referential atomism.[11]

Mental terms do not represent objects simpliciter, they carry bits of information – what a mental term is most likely to represent, then, is what the most likely source of all that information, taken together, happens to be. Given information T about some object S, my term R represents S (and not some similar object P) if it is more likely that the information represented in R corresponds to properties of S than properties of P. R represents S if $P(T|R_S) > P(T|R_P)$. In other words, I have a representation of a cat when the atomistic information carried in my mental term – having pointed ears, night vision, whiskers, etc. – is more likely to represent a cat than a dog.

It is clear now that Usher's account of mental misrepresentation does not hinder my argument. The CTer who would bring up his revision to mental representation wants to say, "Look, mental terms don't have to refer to their causes directly!" Yet Usher's reworking of how mental representation works is, in fact, a softer version of NTDP. Much as Usher says the content of my mental representations depends on my bits of atomistic information leaning more in favor of some S than another, NTDP says that I need to build a theory for any S before I can detect it. Eventually, any S-theory I build contains the same kind of atomistic information that Usher's view on probabilistic representation requires. He reaches the same conclusion I support: causally-created mental representations lack directly external content. So while Usher's reworking evades the worry that CT needs to directly token terms that refer externally, it does not alleviate CT's problems. CT-cum-NTDP still entails LOT – and while a CT with a probabilistic representationalism might not need external content, the LOT framework still does.

[11] Usher, "A Statistical Referential Theory of Content," 317.

4.3. OBJECTION 3

I mention previously that referrers ought in principle be able to be tokened purely ostensively. So, the reader with this objection responds to my argument thus: it is reasonable to accept that under normal circumstances CT requires me to know about a term in order to token it; but, there is still a way to token referrers, namely, through ostensive definitions. I can simply point to a beech and say, "this here is what I mean by 'beech'," still without adequately knowing how to detect beeches. Referrers are still available to CT. So, CT still explains mental content.

I think that this is actually a very legitimate reply. This objection still meets the entire criterion for what I've argued mental referrers are, especially the crucial point about an actual, external beech causing it to be the case that my mental term |beech| has content. It also still meets CT's requirements, since the creation of the term is causally linked to an external beech. Likewise, since the term was tokened through ostensive definition and not detection proper, I cannot say that it is a kind of beech-theory-indicator. The term refers to nothing but the beech I just pointed at. So indeed it seems that we have a genuine beech referrer.

But this reply is not quite right, either. We will have obtained a beech referrer for *that particular beech over there*. But we will not have a beech referrer. The CTer wants to say that whenever I token |beech|, it refers to all beeches; he wants to say that my term |beech| is the same every time. But that is not possible if its content-fixing happens by way of ostensive definition. If the only way to token a direct referrer is through ostensive definition, I do not end up with one referring primitive term – one |beech| or |fox| or |white rat| – I end up with several hundred thousand. The terms whose content I fix from beech-theory-satisfaction do not refer to any external beeches, but the ones whose content I fix ostensively do? Fair enough. But the CTer with this objection ends up having to say that every time I see a beech, my CT and LOT must churn out a new mental term for the particular beech in question. In other words, in wanting to say that CT can token a |beech| that directly refers, this CTer ends up saying that I never actually get |beech|, I only ever get |*this* beech| and |*that* beech|.* This should be enormously unpalatable for any causal theorist.

---

* Dennis Stampe (1997) purports to solve this problem (known in the literature as the 'problem of singularity'). Issues arise with his account, however, regarding the content of terms not tokened causally (in idle thought, for example) and especially regarding mental misrepresentation (Usher, 2001). His argument entails that misrepresentation actually never happens – while clearly, it does. The problem of singularity is outside the main focus of this essay; likewise, for brevity's sake, I refrain from engaging his argument here.

5. FINAL THOUGHTS

Finally, it might be said that eliminating CT leaves no other opportunity for the content of mental terms to be fixed empirically. For what it's worth, however, I admit that I am an empiricist regarding mental content myself. CT certainly is not the only empiricist theory of mental content-fixing. Other empiricist theories of mental content may not have the same consequences – i.e. the LOT hypothesis – that CT itself does. One example is J.J Prinz's (2006) "proxytype theory."* Where CT tokens pseudo-linguistic mental terms, Prinz's theory gets mental models, making mental representation a sort of simulation process. His is a clearly empirical account of mental representation that differs quite fundamentally from the orthodox LOT account that "thinking occurs in a symbolic medium, whose representations have subject-predicate structure and are manipulated by logical rules."[12] Another is Dan Ryder's SINBAD neurosemantics, where the networks of human neural connections learn, through trial and error, to 'model' the world non-referentially.** Empiricism is certainly still available.

But CT is one particular theory of mental content that, upon close inspection, refutes itself. The expectations that CT sets for what the content of its own terms is end up being far above its capacity to fulfill. From the seemingly *prima facie* acceptable CT, the causal theorist builds a LOT hypothesis to comfortably sit in. However, the constraints – namely, NTDP – on the sort of terms whose content CT can account for actually preclude the satisfaction of that LOT framework's promises. It has been shown that, contrary to what LOT expects and what CT itself actually declares, I cannot fix the content of terms as a direct result of external detection. And so I cannot have mental terms that have the sort of content CT and LOT want. The very thing that allows my mental terms to be consistently meaningful in fact omits the possibility that CT provides that meaningfulness. Ultimately, CT fails as a theory of mental content.

---

* Prinz's account draws a divide between primitive concepts themselves (i.e., those with empirical basis) and their various proxytypes, so-called because each stands in as a proxy for the categories and concepts they represent.

[12] Jesse J. Prinz, *Furnishing the Mind: Concepts and Their Perceptual Basis* (Cambridge, MA: MIT Press, 2006), 151.

** Cf. Dan Ryder, "SINBAD Neurosemantics: A Theory of Mental Representation," *Mind & Language* 19 (2004): 211-240.

WORKS CITED

Bridges, Jason. "Davidson's transcendental externalism." *Philosophy and Phenomenological Research*, 73.2(2006): 290–315.

Burge, Tyler. "Individualism and Psychology." *Philosophical Review*, 95(1986): 3–45.

Cummins, Robert. "The LOT of the Causal Theory of Mental Content." *Journal of Philosophy*, 94(1997): 535–542.

Falvey, Kevin and Joseph Owens. "Externalism, self-knowledge, and skepticism." *Philosophical Review*, 103(1994): 107–37.

Fodor, Jerry. *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press, 1990.

Noordhof, Paul. "Environment-Dependent Content and the Virtues of Causal Explanation." *Synthese*, 149(2006): 551–575.

Prinz, Jesse J. *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: MIT Press, 2006.

Usher, Marius. "A Statistical Referential Theory of Content: Using Information Theory to Account for Misrepresentation." *Mind and Language*, 16(2001): 311–334.

Ryder, Dan. "SINBAD Neurosemantics: A Theory of Mental Representation." *Mind & Language*, 19(2004): 211–240.

Rupert, Robert. "Coining Terms in the Language of Thought: Innateness, Emergence, and the Lot of Cummins's Argument against the Causal Theory of Mental Content," *Journal of Philosophy*, 98(2001): 499–530.

Speaks, Jeff. "Is Mental Content Prior to Linguistic Meaning?" *Noûs*, 40(2006): 428–67.

On the Interpretation of Marx's Critique of Hegelian Metaphysics

Luis Salazar
Florida International University

Karl Marx's thought is often interpreted as intentionally and openly contradicting Hegel's metaphysics (i.e., Hegel's theory of the nature of reality).* That is to say, Marx is taken as straightforwardly *rejecting* Hegel's metaphysics. The question of importance for this particular project is: how does Marx understand Hegel's metaphysics? If one assumes that Marx rejects Hegel's metaphysics, then it seems quite easy to assume that Marx understands Hegel to believe in a *false* metaphysical system, even though he may find brilliance in, and praise Hegel for, other aspects of his thought. For example, Frederick Beiser writes that one objection to Hegel's Theory of History is taken from "Marx's *attack* upon Hegel in the *German Ideology*. According to Marx, Hegel's historicism is vitiated by his metaphysics. Hegel did not go far enough in attempting to transform metaphysics; the point is to abolish it."[1] The aim of this paper is to provide an argument in order to clarify and push back precisely this interpretation of Marx's understanding of Hegel's metaphysics.

To begin, I will briefly set the stage by discussing the common interpretation of Marx's understanding of Hegel's metaphysics and his supposed stance towards it. I will proceed to argue that this common interpretation is actually a misinterpretation. This will be done primarily through an analysis of the concluding section of the *Economic and Philosophic Manuscripts of 1844* titled, "Critique of the Hegelian Dialectic and Philosophy as a Whole."[2] Of course, this will be done with a proper understanding of what Marx understands as "critique." Hence, during the course of this analysis a moment will be taken to consider what Marx understood himself to be doing when he engaged in a critique. Through this inquiry, I hope to show that Marx does *not* understand himself to be rejecting Hegel's metaphysics; at best, Marx's rejection only applies to what he considered improper or cowardly interpretations of Hegel's philosophy by Marx's contemporaries.

To be clear, this argument will not question whether Marx's interpretation and understanding of Hegel is accurate or not (i.e., whether Marx understands Hegel as he *should* be understood). Consequently, the argument will also *not* take the form of discussing how Hegel's metaphysics should be understood and then attempting to show how Marx's thought coincides with that understanding.

---

* The term contradiction I use solely in the sense that its Latin roots suggest *contra dicere* or "to speak against". It must be clear that by "Hegel's Metaphysics," I am not referring to his metaphysics as he himself understood them, or as they are understood by Hegel scholars, Marxists, or people outside the philosophical discipline. I am referring simply to Hegel's Metaphysics *as Marx understood them*. Thus, as will be seen, this essay questions whether Marx understands himself as opposing what he understands to be Hegel's metaphysics.

[1] Frederick C. Beiser, "Hegel's Historicism," in *The Cambridge Companion to Hegel*, ed. Frederick C. Beiser (Cambridge: Cambridge UP, 1993), 270-300. First italics added. Although I do not use *The German Ideology* as a major part in this analysis, I hold that the argument in this paper applies to the critique Marx offers in *The German Ideology* as well.

[2] Karl Marx, "Economic and Philosophic Manuscripts of 1844," in *The Marx-Engels Reader: Second Edition*, ed. Robert C. Tucker (New York: W.W. Norton & Company, 1978), 277.

Rather, Hegel's metaphysics will only be looked at from Marx's understanding (i.e., dependent on Marx's opinions and his interpretation of Hegel's thought).

The immense influence of Hegel over Marx is, of course, no secret. Neither is it a secret that much of Marx's own thought seems to be direct appropriation, or alteration, of Hegel's. For example, Beiser writes, "Hegel anticipates much in Marx's own materialist historicism. Marx's materialism is indeed little more than Hegel's historicism without the metaphysics of the absolute idea."[3] Furthermore, in his paper *Hegel and Marxism*, Allen Wood gives an extensive account of where the theories of Marx and Hegel coincide. Part of the aim of that paper was to show that, "...Marx's conception of proletarian revolutionary practice... is a conception of self-transparent [i.e., freedom, or full consciousness of what one does, in a social context] world historical agency that is consistent with the strictures of a Hegelian philosophy of history."[4] Several other writers may be quoted expressing a similar relation between Hegel and Marx. What is important to note is that while different authors may give more or less similar explanations for the influence of Hegel over Marx, they seem to agree that both thinkers are at odds with each other on a particular issue. That issue is Hegel's *idealistic* metaphysics, or the idea that world history is governed by *Spirit*, rather than objective human *beings*.

There must, of course, be a source that attributes this position to Marx. In short, it can be characterized through a quote from *Capital*:

> My dialectical method is not only different from the Hegelian, but is its direct opposite. To Hegel, the life-process of the human brain, i.e., the process of thinking, which, under the name of "the Idea," he even transforms into an independent subject, is the demiurgos of the real world, and the real world is only the external, phenomenal form of "the Idea." With me, on the contrary, the ideal is nothing else than the material world reflected by the human mind, and translated into the forms of thought.[5]

In other words, it seems that Hegel holds that everything that is, is Idea. "Rather than seeing the basic determining force of history as the concrete material needs of particular human beings, Hegel claims that they are found in the absolute idea. Allegedly, he holds that 'the world is ruled by ideas, that

---

[3] Beiser, *Hegel's Historicism*, 278-279.

[4] Allen Wood, "Hegel and Marxism," in *The Cambridge Companion to Hegel*, ed. Frederick C. Beiser (Cambridge: Cambridge UP, 1993), 441.

[5] Karl Marx, *Capital, Volume 1*, trans. Samuel Moore and Edward Aveling (New York: Random House Inc., 1906), 25.

ideas and concepts are its determining principles.'"[6] There is no need to go into great detail here to the specific details of Hegel's metaphysics for reasons stated above, but it is important to note the relation between Marx and Hegel, exhibited here by Marx himself, as a relation of opposition. Furthermore, Marx himself having expressed an understanding of Hegel as an idealist may seem to undermine precisely what this paper attempts to show, thus, an explanation is due as to why this is not the case.

One explanation would be to note that what Marx is comparing are dialectical *methods*. The quote above is taken from the *Afterword to the Second German Edition* of *Capital* where he is addressing how "little understood" *his own method* in *Capital* had been. Marx explains his method as that of inquiring into the subject-matter to find a pattern or form of development (movement). Once that pattern of development could be described adequately enough, then it could be used give a description of the phenomena as if it were a law (e.g., describing how X has gone about and how it should continue to do so). That is, the material conditions are examined and then the ideal (theory) arises from proper reflection of the particular material conditions. He contrasts his dialectical method from that of Hegel by claiming that Hegel proceeds in the opposite direction (i.e., Hegel begins at the Idea which then takes the resulting external phenomena as itself manifested outside itself).[7] Clearly, anyone familiar with the writings of both thinkers will immediately agree with this distinction. Further, while it is quite clear that Marx is rejecting the Hegelian dialectical method (i.e., one can only safely assume what was quoted above to be making reference to Hegel's methodology of inquiring on a subject-matter and his manner of presenting it), it is still far from clear that Hegel's dialectical method is one and the same with his metaphysics. That is to say, that Marx is rejecting Hegel's idealism *as a method* is clear, but it must then be the case that Marx understood Hegel's metaphysics as one and the same as Hegel's dialectical method, in order to claim that Marx rejected Hegel's metaphysics. Thus, closer inspection is due to see what can be said of this situation.

In *Capital*, Marx goes on to claim that he had *criticized* the "mystifying side of Hegelian dialectic" nearly thirty years before.[8] This reference that Marx offers should not be taken lightly. The claim in itself does not seem to offer much. However, a closer look should be given to Marx's earlier writings in order to properly assess whether Marx was also referring to Hegel's metaphysics in his rejection of Hegel's dialectical method. The fact that Marx at least vaguely refers the reader to his previous work on this issue is significant. It suggests strongly

---

[6] Wesser, *Hegel's Historicism*, 278.
[7] Marx, *Capital, Volume I*, 24-25.
[8] Ibid., 25.

that whatever Marx concluded in his critique of Hegel, written in his twenties, was still valid in his fifties. If Marx indeed opposes Hegel's metaphysics, then the conclusive proof should be found in these earlier, more philosophical, writings.* Thus, we now turn to this critique itself to inquire if, in fact, the critique does strengthen and support the conviction that Marx opposes what he understands to be Hegel's metaphysics. As I will show, the critique does *not* support that conviction. What it will demonstrate, however, is a rather intricate understanding, on Marx's part, of Hegel, which will shed new light on and give depth to the distinction Marx makes between his own dialectical method and Hegel's.

From this point forth, this demonstration will proceed by taking a closer look at Marx's understanding of Hegel's dialectical method and Hegel's metaphysics primarily through an analysis of the concluding chapter of the *Economic and Philosophic Manuscripts of 1844*; all the while, attempting to understand the meaning of Marx's criticism of the "mystifying side of Hegelian dialectic."** It is this text which may provide the greatest detail and, consequently, the greatest insight into interpreting how Marx understands both. However, prior to this, one important issue *must* be addressed.

The concluding chapter of the *Economic and Philosophic Manuscripts of 1844* is interestingly titled, "Critique of the Hegelian Dialectic and Philosophy as a Whole." Thus, it will be assumed that Marx is careful with the usage of the term "critique," that is to say, there is a specific activity that Marx understands himself to be engaged in while critiquing. Hence, prior to attempting to derive anything from the critique itself it is of utmost importance to have a clear understanding of what Marx means by "critique." One must be careful with the manner in which the word critique (from the German *Kritik*) is interpreted here, so that it not be taken strictly in a negative sense (i.e., as a sort of fault-finding endeavor). The German word *Kritik* does not have this connotation, although faults in the object of critique may derive as a result of performing a critique.

Fortunately, Marx himself directly addresses the concept of criticism in a letter he wrote to Arnold Ruge in 1843. Although Marx is not as clear as one would like, he does describe the goal of criticism in various manners within this letter. For example, when describing the aim of the journal, *Deutsch-*

---

* We will put aside the praises Marx offers Hegel within the afterword to *Capital* because they will offer nothing (for the moment at least) in the attempt to comprehend Marx's understanding of Hegel's metaphysics; i.e., the positive remarks will lead no closer to establishing what exactly Marx may have meant when he claimed that he had criticized the "mystifying side of Hegelian dialectic."

** Interestingly enough, the *Economic and Philosophic Manuscripts* as well as the *Contribution to the Critique of Hegel's Philosophy of Right* were unpublished during Marx's lifetime. Hence, Marx, in *Capital*, is directing the person reading his work during his own lifetime to something he/she probably would not have been familiar with. Perhaps a very select group of people, i.e., those who Marx allowed access to these particular works, were aware of exactly what Marx was referencing.

Französische Jahrbücher, as a *critical* journal, Marx writes, "so, we can express the trend of our journal in one word: the work of our time to clarify to itself (critical philosophy) the meaning of its own struggle and its own desires." From this, criticism can be described as an endeavor to make clear, understandable, what is already there but in an unclear or mystical form; "we only show the world what it is fighting for, and consciousness is something that the world *must* acquire, like it or not." Moreover, there is also a sense in which it implies "bringing out" whatever is hidden, whether consciously or unconsciously, regardless of consequences for doing so — "the criticism must not be afraid of its own conclusions, nor of conflict with the powers that be."[9] It is important, then, to keep in mind the sense in which a critique is an endeavor not only to make clear that which is unclear, but to make clear and bring to light that which was purposefully made unclear.* Marx, then, in critiquing Hegel, is not necessarily attempting to point out the faults in Hegel's system. Rather, he is attempting to clarify Hegel's work, to do away with the mystical aspect with which, Marx believed, Hegel disguised, or hid, something other within his system.

Hence, this understanding of "critique" almost immediately sheds light on what can be expected within the texts that Marx refers to as critiques, especially those which concern Hegel. For example, Marx writes in the preface to the *Economic and Philosophic Manuscripts of 1844*, "in contrast to the *critical theologians* of our day, I have deemed the concluding chapter of the present work — the settling of accounts with *Hegelian dialectic* and Hegelian philosophy as a whole — to be absolutely necessary, a task not yet performed."[10] If one then takes into account how Marx understands "critique," one cannot help but conclude from the very preface that Marx seems to be claiming that there is, within Hegel's dialectical method, something that has not yet been made clear — something hidden. Of course, the use of the word "hidden" implies that the mysticism was by *design*, which is an assumption that must be made if one takes Marx seriously as a thinker; otherwise, one could simply claim that Marx was drawing arbitrarily from Hegel's texts. However, Marx's manner of treating the subject, as will be shown, is clearly one in which he understands himself to have understood Hegel to have purposefully made his writings obscure.

Whatever it may be that is concealed within, or by, Hegel's dialectical method, is clearly something that Marx wants to make explicit. Reading through the remainder of the paragraph only strengthens this conclusion:

---

[9] Karl Marx, "For a Ruthless Criticism of Everything Existing," in *The Marx-Engels Reader: Second Edition*, ed. Robert C. Tucker (New York: W.W. Norton & Company, 1978), 13-15.

* Other quotes which express manners in which Marx understood the goals of critique include: "bringing out the true significance underlying this system," "analyzing the mystical consciousness, the consciousness which is unclear to itself," as well as, "it is a matter of *confession*, no more."

[10] Marx, *1844 Manuscripts*, 68.

> This *lack of thoroughness* is not accidental, since even the *critical* theologian remains a *theologian*. Hence, either he had to start from certain presuppositions of philosophy accepted as authoritative; or if in the process of criticism and as a result of other people's discoveries doubts about these philosophical presuppositions have arisen in him, he abandons them without vindication and in a cowardly fashion, *abstracts* from them showing his servile dependence on these presuppositions and his resentment at this dependence merely in a negative, unconscious and sophistical manner.[11]

Marx is claiming that the Hegelian dialectic, and Hegel's philosophy, are taken by some as authoritative and never questioned, or if by chance they are brought into question, then they are rejected, abandoned.* That is to say, some of these "critical theologians" simply swallow up the mysticism, while the others that may come to *discoveries* that cast doubt on the mysticism simply continue to support the mysticism by abandoning it or rejecting it completely. They recognize a dependence on the mysticism and yet are angered by this dependence; however, they do not work against it by trying to clarify it.** Marx's claim is then that while the "critical theologians" may have become aware of what Marx believes is the actual purpose of the mysticism, they have decided to maintain and, to some extent, support the mysticism (i.e., they have strongly refused to make clear that which is mystical in Hegel), either willingly or unwillingly, rather than clarify it.

It must be admitted that what was quoted above, if not looked at carefully, may lend itself to support the position that Marx strongly rejected Hegel's metaphysics. Although this concern cannot be fully addressed as of yet, it should be noted that it is *not* the case that Marx is addressing Hegel directly. That is, what are clearly attacks on Marx's part are towards the "critical theologians" of his day, or those who have interpreted Hegel in a certain manner that fits the descriptions above. Marx is explaining that his critique of the Hegelian dialectic and Hegel's philosophy is regarding an issue that he and his contemporaries find themselves at odds about. Thus, what can be said is that Marx is attacking the decision, made by his contemporaries, to refrain

---

[11] Marx, *1844 Manuscripts*, 68.

* Note: to question here is not to reject or find faults, but rather to inquire into what is really there, to clarify.

** Marx further writes, "in this connection the critical theologian is either forever repeating assurances about the *purity* of his own criticism, or tries to make it seem as though all that was left for criticism to deal with now was some other immature form of criticism outside itself—say eighteenth-century criticism—and the backwardness of the *masses*, in order to divert the observer's attention as well as his own from the *necessary* task of settling accounts between *criticism* and its point of origin—Hegelian *dialectic* and German philosophy as a whole—from this necessary raising of modern criticism above its own limitation and crudity" (Marx, *1844 Manuscripts*, 68-69).

from clarifying that which takes a mystical form in Hegel's writings.*

This then leads us to ask: What is it that Marx felt that he was at odds with his contemporaries about? What is it that Marx, in his early writings, is attempting to make clear? In order to attempt to answer this, it is appropriate to inquire directly into Marx's critique of Hegel. In short, the "Critique of the Hegelian Dialectic and Philosophy as a Whole" is, up until the last few paragraphs, nothing more than exegesis of the final section of Hegel's *Phenomenology of Spirit*, which Marx describes as "the true point of origin and the secret of the Hegelian philosophy."[12] If one abides by the rule that a text should be judged for what it reads, that is, for what is *there* (and rightfully so), then there should be ***absolutely*** no reason to believe from reading this text that Marx's intent is the rejection of Hegel's metaphysics. This, of course, requires clarification.

The discussion up until this point implies that for Marx, there is a potential difference between Hegel's dialectical method and Hegel's metaphysics. Importantly, from what has been discussed so far, it seems safe to assume that Hegel's dialectical method can also be roughly described as Hegel's method of writing and presenting his ideas. That being said, it is only proper to admit that it is indeed the case that Marx rejects Hegel's dialectical method, i.e., *presenting*, or *writing*, in the manner Hegel does. It is precisely this which Marx is attacking his contemporaries for doing, as is evident from Marx's attack on Bauer: "such expressions do not even show any verbal divergence from the Hegelian *approach*, but on the contrary, repeat it word for word."[13] The use of the word 'approach' should be a clue. Marx is attacking Bauer for his continued use of Hegel's dialectical method when, in Marx's opinion, it is unnecessary and cowardly to do so if, indeed, Bauer grasps that which is concealed by the method. In other words, Marx is not explicitly rejecting Hegel's metaphysics, but because the attack is directed towards his contemporaries, what *is* clear is that he rejects the continued use of Hegel's dialectical method.

A further point remains to be addressed, and that is the question of how Marx understood Hegel's metaphysics. This paper aims to show that Marx

---

* Along the same lines, in *The German Ideology* Marx writes, "German criticism has, right up to its latest efforts, never quitted the realm of philosophy. Far from examining its general philosophic premises, the whole body of its inquiries has actually sprung from the soil of a definite philosophical system, that of Hegel. Not only in their answers but in their very questions there was a mystification. This dependence on Hegel is the reason why not one of these modern critics has even attempted a comprehensive criticism of the Hegelian system, however much each professes to have advanced beyond Hegel. Their polemics against Hegel and against one another are confined to this—each extracts one side of the Hegelian system and turns this against the whole system as well as against the sides extracted by the others" (Karl Marx, "The German Ideology," in *The Marx-Engels Reader: Second Edition*, ed. Robert C. Tucker (New York: W.W. Norton & Company, 1978), 148.) Again, it is unclear here whether the attack is to any significant degree against Hegel; however, what is clear is Marx's attack against other German critics of Hegel. Marx attacks *their* mysticism, as well as, their inability (or unwillingness) to understand Hegel's philosophy as a whole.

[12] Marx, *1844 Manuscripts*, 109.

[13] Ibid, 106.

understands Hegel to have purposefully presented his work in the manner he did in order to conceal something. Showing that Marx did not understand Hegel's metaphysics as he understood Hegel's dialectical method is, then, the aim of the remainder of this paper. Furthermore, it is not within the scope of this paper to give a detailed analysis of Marx's understanding of Hegel's metaphysics, especially when a brief overview will suffice. This should bring into question the common interpretation, where Marx is understood as believing Hegel to actually be an idealist, and further, rejecting and conflicting with Hegel's thought because of this supposed idealism.

What Marx presents in his critique of the Hegelian dialectic is precisely an inquiry into the very limitations of Hegel's dialectical method. According to Marx,

> the *Phenomenology* is...an occult critique—still to itself obscure and mystifying criticism; but inasmuch as it keeps steadily in view man's *estrangement*, even though man appears only in the shape of mind, there lie concealed in it *all* the elements of criticism, already *prepared* and *elaborated* in a manner often rising far above the Hegelian standpoint.[14]

This quote alone embodies the aim of the entire critique Marx is presenting in the final chapter of the *Economic and Philosophic Manuscripts of 1844*. Marx presents from the start what he describes as a "double error in Hegel," but this *error* is not to be interpreted as a claim that Hegel is incorrect in any specific way, nor should the rest of the section be interpreted as an attempt by Marx to show the error in Hegel's ways; taking this route would, in fact, be a harsh deviation from what Marx had until then been presenting as the purpose of critiquing Hegel's philosophy in the first place.[15*] The "double error" is nothing more than an inconsistency, or a contradiction, that Marx interprets Hegel himself to have *set up* (or as quoted above *prepared* and *elaborated*) to make a point that is not explicit.

The "Critique of the Hegelian Dialectic and Philosophy as a Whole" is thus a supposedly un-mystified version, presented by Marx, of what he believes to be Hegel's own "occult" critique.** To elaborate, Marx, by concentrating solely

---

[14] Marx, *1844 Manuscripts*, 111.

[15] Ibid., 110.

* That is to say, it would turn away from the question of how to interpret Hegel, which until this point had been the only dilemma recognized between himself and his contemporaries that Marx had been dealing with within the *1844 Manuscripts*.

** That is, to show all the elements of criticism implicit in Hegel, and how they are "*prepared* and *elaborated* in a manner often rising far above the Hegelian standpoint," or, put in other words, to show how Hegel *designs* his writings to go beyond the idealistic standpoint (approach) from which he presents them.

on the final section of the *Phenomenology of Spirit* and by later bringing in other aspects of Hegel's philosophy (e.g., Hegel's *Encyclopedia, Logic*), attempts to demonstrate how this hidden critique takes form within Hegel's writings. By analyzing the *Phenomenology* and Hegel's other works, Marx argues that Hegel, while writing in an idealistic manner (i.e., from the standpoint of *mind*), leaves enough of the tools necessary to see beyond the idealism. He argues that Hegel's position is rather a "consistent naturalism or humanism" which "distinguishes itself from both idealism and materialism, constituting at the same time the unifying truth of both."[16] That is to say, the abstract point of view (i.e., that of mind, thought), *as well as* the materialistic point of view, both ultimately undermine themselves when considered as distinct and apart from one another.

Regarding the idealistic view, Marx writes in a footnote:

> [W]hat Hegel does is to put in place of these fixed abstractions the act of abstraction which revolves in its own circle. In doing so, he has the merit, in the first place, of having indicated the source of all these inappropriate concepts which, as originally presented, belonged to disparate philosophies; of having brought them together; and of having created the entire compass of abstraction exhaustively set out as the object of criticism, instead of some specific abstraction.[17]

Thus, according to Marx, Hegel's intention is not really to present some specific type of abstraction above all others. Rather, with regard to abstract positions, Hegel laid out, specifically, what was to be critiqued and superseded.

Further, regarding Nature or the materialistic view, Marx consults Hegel's *Philosophy of Nature* and concludes that what Hegel presents there is Nature in-itself. Thus:

> nature *as nature*—that is to say, in so far as it is still sensuously distinguished from that secret sense hidden within it—nature isolated, distinguished from these abstractions is *nothing*—a nothing *proving itself to be nothing*—is *devoid of sense*, or has only the sense of being an externality which has to be annulled.[18]

This is a nature distinct from that which is understood by a naturalism

---

[16] Marx, *1844 Manuscripts*, 115.
[17] Ibid., 123. (Footnote 5).
[18] Ibid., 124.

or humanism, which Marx attributes to Hegel. Just as considering the world as thought in-itself is an error, considering nature abstractly, or as nature in-itself, is also an error — it is "*devoid of sense.*" It results in considering Nature as "a mistake, a defect, which ought not to be."[19] It is in this manner that Marx understands Hegel to argue against materialism. That is to say, according to Marx, Hegel is not rejecting nature with respect to the human being, but rejecting abstract nature — nature in-itself. It is this more complex, intricate understanding of Hegel which should be taken as Marx's understanding of Hegel's metaphysics. However, this metaphysical position is not what Marx rejected in *Capital.*

Taking Hegel's dialectical method to be the same as Hegel's actual stance regarding the nature of reality (i.e., equating Hegel's method of presenting his written work with his metaphysics) is a misunderstanding that easily earns Hegel the title of idealist, and earns Marx a position opposed to Hegel's. Yet, close inspection of Marx's texts seems to indicate that Marx does not understand Hegel's metaphysics and Hegel's dialectical method as identical things. For Hegel, according to Marx, the Absolute Idea is nothing more than "abstract thinking that gives itself up and resolves on *intuition.*" Marx continues,

> ...the abstract thinker learns in his intuition of nature that the entities that he thought to create from nothing, from pure abstraction—the entities he believed he was producing in the divine dialect as pure products of the labour of thought forever weaving in itself and never looking outward—are nothing else but *abstractions* from *characteristics of nature.*"[20*]

The same point is made in Marx's "Contribution to the Critique of Hegel's Philosophy of Right," where he writes:

> Empirical actuality is thus accepted as it is. It is also expressed as rational, but it is not rational on account of its own reason, but because the empirical fact in its empirical existence has a different significance from it itself. The fact which is taken as a point of departure is not conceived as such, but as a mystical result. The actual becomes a phenomenon. Nor has the idea any other purpose than the logical one of being "explicitly infinite actual mind."[21]

---

[19] Marx, *1844 Manuscripts*, 125.

[20] Ibid., 122-24.

[*] Intuition meaning "to be aware through the senses." See footnote 4 in *Critique of Hegelian Dialectic.*

[21] Karl Marx, "Contribution to the Critique of Hegel's *Philosophy of Right*," in *The Marx-Engels Reader: Second Edition*, ed. Robert C. Tucker (New York: W.W. Norton & Company, 1978),17-18.

That is to say, Marx holds that Hegel's metaphysics, properly understood, has as a starting point some material condition. It is the methodology Hegel employs which does not allow for this "point of departure" to be expressed. Hence, what Marx claims his own dialectical method to be in opposition to is *not* what Marx himself understands to be Hegel's metaphysics, but rather, what Marx understands as Hegel's dialectical method.

This discussion gives rise to various other important questions which do not fall within the scope of this paper. However, what is of importance to this paper is the question of Marx's understanding of Hegel's metaphysics, and whether or not he rejected it. The answer has turned out to be quite complex; it has required a proper understanding of what Marx means by critique, that is, that Marx's aim in a critique is clarification of something unclear, mystical, hidden. Moreover, analysis of the final chapter of the *Economic and Philosophic Manuscripts of 1844* has shown what Marx actually meant, in his later years, when he claimed that he had "critiqued" Hegel's mysticism. That is, Marx understood himself to have clarified, de-mystified, what Hegel had hidden within his writings. What is important about this realization is that the very fact that Marx differentiates between Hegel's dialectical method, and his metaphysics, shows that his claim in the afterword to *Capital* is not a rejection of *Hegel's metaphysics*. Rather, it shows that Marx rejected the further use of *Hegel's dialectical method*, and in particular, its use by his contemporaries who continued to support the method. It is these interpreters that Marx deems cowardly, if in fact they grasp what Hegel hides with his mysticism. Taking all of this into account, the result is a renewed understanding of what Marx meant when he claimed that he critiqued Hegel's mysticism — an understanding that potentially does away with the notion that Marx understood himself to have gone beyond Hegel in any significant manner, and brings to the fore the importance of Hegelian thought within Marx's philosophical system.

BIBLIOGRAPHY

Beiser, Frederick C. "Hegel's Historicism." In *The Cambridge Companion to Hegel*. Edited by Frederick C. Beiser. Cambridge: Cambridge UP, 1993. 270-300.

Marx, Karl. *Capital, Volume 1*. Translated by Samuel Moore and Edward Aveling. New York: Random House Inc., 1906.

Marx, Karl. "Contribution to the Critique of Hegel's *Philosophy of Right*." In *The Marx-Engels Reader: Second Edition*. Edited by Robert C. Tucker. New York: W.W. Norton & Company, 1978. 16-25.

Marx, Karl. "Economic and Philosophic Manuscripts of 1844." In *The Marx-Engels Reader: Second Edition*. Edited by Robert C. Tucker. New York: W.W. Norton & Company, 1978. 66-125.

Marx, Karl. "For a Ruthless Criticism of Everything Existing." In *The Marx-Engels Reader: Second Edition*. Edited by Robert C. Tucker. New York: W.W. Norton & Company, 1978. 12-15.

Marx, Karl. "The German Ideology." In *The Marx-Engels Reader: Second Edition*. Edited by Robert C. Tucker. New York: W.W. Norton & Company, 1978. 146-200.

Wood, Allen. "Hegel and Marxism." In *The Cambridge Companion to Hegel*. Edited by Frederick C. Beiser. Cambridge: Cambridge UP, 1993. 414-444.

The Constituents of Phenomenal Experience

Corey Cusimano
University of North Carolina, Chapel Hill

INTRODUCTION

Representationalism, as a theory of phenomenal consciousness, is the theory to beat. It holds that "the mind has no distinctive properties that outrun its function and intentional properties."[1] This view has been most recently articulated by Lycan,[2] Tye,[3] Dretske,[4] Harman,[5] and Carruthers.[6] Their main opponents are *Phenomenists*.[7] Phenomenists hold that the mind has properties besides merely the intentional (or functional) that play a constitutive role in the nature of our phenomenal experience.[8] The debate between these two camps usually takes the form of phenomenists showing that our phenomenal experiences *outrun* the functional and intentional content available; however, most recently Block has argued, instead, that the representational content of perception is "incompatible with the phenomenal character of perception."[9] The purpose of this paper is to examine the various competing theories from the perspective of Block's historical challenges to representationalism. The goal is to put his latest article, "Attention and Mental Paint," in perspective of the larger debate mentioned above.

When discussing phenomenal consciousness, it is important to distinguish between several different (though often bundled) concepts. Though the two are often combined, we can separate qualitative features of mental states, "qualia," from the related concept "what it is like."[10] Qualia are ostensibly properties of the world while "what it is like" is, roughly, the character of experiencing a quale. Qualia are what the world seems like (e.g., the apparent color of a physical object in the world). An easy heuristic for distinguishing the two is that qualia are describable in ordinary English words like "red" or "square," while "what it is like" is ineffable.[11] Block uses the term "qualia" to include all aspects of *phenomenist* phenomenal consciousness (therefore, by stipulation, phenomenal properties of consciousness that outrun the intentional, functional, and cognitive).[12] I mention this only to say that when I use "qualia," I am not using it in this sense.

---

1 William G. Lycan, *Consciousness and Experience* (Cambridge, MA: MIT Press, 1996), 67.

2 William G. Lycan, "In Defense of the Representational Theory of Qualia (Replies to Neander, Rey, and Tye)," *Nous*, 32(1998), 479-87.

3 Michael Tye, *Consciousness, Color, and Content* (Cambridge, MA: MIT Press, 2000), 69-95.

4 Fred Dretske, *Naturalizing the Mind* (Cambridge MA: MIT Press, 1995), 1-39.

5 Gilbert Harman, *Reasoning, Meaning, and Mind* (Oxford: Oxford University Press, 1999), 235-76.

6 Peter Carruthers, *Phenomenal Consciousness: A Naturalistic Theory* (Cambridge: Cambridge University Press, 2000), 122-27.

7 Ned Block, "Mental Paint and Mental Latex," *Philosophical Issues* 7(1996): 19.

8 Ned Block, "Mental Paint," in *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, ed. M. Hahn and B. Ramberg (Cambridge MA: MIT Press, 2003), 165-200.

9 Ned Block, "Attention and Mental Paint," *Philosophical Issues* 20 (2010): 25.

10 Carruthers, *Phenomenal Consciousness*, 122-27.

11 William G. Lycan, "Representational Theories of Consciousness," The Stanford Encyclopedia of Philosophy, ed. Edward N. Zalta, <http://plato.stanford.edu/archives/fall2008/entries/consciousness-representational/>.

12 Block, "Mental Paint," 165-200.

A quale, in the sense defined above, can be a variety of things: the color of an object (or an after-image); the pitch, volume, or timbre of a sound; a taste, texture, and so on, for the different properties we seem to "sense" in the world. What constitutes these qualia is what is at issue. Before jumping into the dialectic, it is worth explicating the variety of views at stake. We can distinguish between pure representationalism, strong representationalism, and weak representationalism.

THE THEORIES

Pure representationalism holds that having intentional content is all that is required for qualia. No one adheres to this view for the simple reason that it would mean that thermometers (and the like, to say nothing of pencil marks on paper) would have phenomenal experiences. Strong representationalism holds that, instead, intentional content of a certain kind results in a quale. "Of a certain kind" is usually specified by materialist or functionalist concerns. This is the view that Dretske, Tye, and Lycan hold. Weak representationalism is the view inhabited by Block (and a few others) who posit that some ontologically "new" property is necessary for phenomenal experience.* Block calls this view "phenomenism" and holds that our experiences have phenomenal properties beyond "qualia" as explicated above.

There are, however, a variety of in-house disputes among the strong representationalists that can be exploited by the dialectic that Block engages in. One of the clearest differences is adoption of the transparency thesis. Tye[13] and Dretske[14] hold a strong transparency thesis while Lycan[15] does not. The transparency thesis (as explicated by Tye and Dretske) holds that the properties we are aware of in our experiences are only those properties out in the world that we perceive.** For a representationalist, this means that the only properties we are aware of are properties of external objects that we are representing. They hold that when we experience something (like the blue on a car), the only features of our experience we can attend to are features of the external world (in this case, *the car's* blueness). Tye and Dretske, after Harman,[16] argue that this is a fact of introspection.

---

* Everyone agrees that qualia are non-conceptual. There are a plethora of good arguments for this view (and some important philosophical conclusions because of it) but, because they play no role in the following exposition, I will not discuss them here.

[13] Tye, *Consciousness, Color, and Content*, 45-54.

[14] Dretske, *Naturalizing the Mind*, 1-39.

[15] Lycan, *Consciousness and Experience*, 109-42.

** Block (2010) attempts to split the transparency thesis into two parts, with a positive claim and a negative claim. More on this below.

[16] Gilbert Harman, "The Intrinsic Quality of Experience," *Philosophical Perspectives* 4(1990): 34-38.

Lycan does not adopt the transparency thesis and argues that qualia are only a part of the "overall feel" of conscious experiences that include "what it is like." Tye and Dretske would say that the "overall feel" is exhausted by the qualia while Lycan would not. Tye and Dretske hold that there is no "what it is like," but only qualia, and they argue that introspection (insomuch as introspection is what leads to the "what it is like" experience) does not exist.* However, the "overall feel" or "what it is like" that Lycan argues for is not restricted to cases of introspection. That is, the "overall feel" can be a property of experience without it being so because of introspection.** Finally, while Tye and Dretske treat transparency as motivation for representationalism, Lycan takes metaphysical concerns (specifically materialism) as a primary motivating factor. These differences lead to different views of what constitutes having a conscious experience (as opposed to an unconscious one). Tye and Dretske hold lower-order views while Lycan holds a higher-order view.

The differences in these views lead to different supervenience claims (and therefore different strategies for rebuttal). Strong-representionalists like Tye and Dretske hold that there cannot be any change in the phenomenal character of an experience without a corresponding change in representational content. I will refer to their views in this paper as *T-Representationalists.*‡ Lycan holds a supervenience claim of the following sort: there cannot be a phenomenal difference without a corresponding change in representational or functional content. I will call a view like Lycan's a *functionalist* view. Block's phenomenist view holds that phenomenal states supervene on representational content and non-representational content like mental paint (or, in places, mental oil or latex). Block, for reasons not discussed so far, eschews functional considerations from playing a constitutive role in phenomenal experience; however, he argues that even if we accept they do, they are still unable to account for the phenomenal difference supposed in a variety of thought experiments.

MENTAL PAINT AND MENTAL OIL

Block argues that the phenomenal qualities that outrun the representational and the function contents of experiences are "mental paint" and "mental oil."[17] Aside from the inverted (or shifted) spectra arguments, Block believes that mental paint and mental oil correspond to qualities of phenomenal experiences found in cross-modal experiences (i.e., visual versus haptic sensations) and

---

* To make this clearer: Tye and Dretske argue that there are no "higher-order" types of experience or consciousness (read: experiences of experiences) while Lycan does.

** Examples include bodily sensations, which will be discussed later.

‡ To stand for, roughly, "transparency representationalists."

17 Block, "Mental Paint," 165-200.

bodily sensations (like pain or orgasms). Block posits the following definitions of mental paint and mental oil:

> Mental Paint: (Mental) properties of the representing experience (i.e., properties of the experience that represent, say, the redness of a tomato).
> Mental Oil: (Mental) properties of the experience that don't represent anything.

The concept of mental paint was introduced by Gilbert Harman as properties of mental experience that were *doing the representing* (as opposed to the properties *being represented*).[18] That is, if the red on a painting represents a tomato, there is still the paint itself (and the properties of the paint) that are parts of the painting. Harman went on to argue that these properties (mental paint) were not a part of phenomenal experience by arguing for the transparency thesis that was explicated above.* Mental paint (as it is described here) is a rejection of the transparency thesis. Mental oil is more elusive but, Block assures us, is the kind of property that inheres in experiences of orgasms (and other bodily sensations).

In relation to our exposition of the theories above, Lycan holds that something like mental paint exists but, rather than being ontologically "new," is explained in terms of functional content. We will turn now to the arguments for mental paint and mental oil.

SENSORY MODALITY AND MENTAL PAINT

Block's dialectical goal in a variety of papers is to show that phenomenal experience "outruns" the supervenience claims made by Tye, Dretske, and Lycan. That is, hold constant the representational content, functional content, and whatever other cognitive considerations you enjoy, and then see if there can be two phenomenally different experiences. This type of strategy has been adopted by a number of authors** and takes the form of alleged counter examples. There is a long history of this form of debate.‡ T-Representationalists typically respond by simply showing that there is, in fact, a representational difference accompanying the phenomenal difference.

One example that Block uses is the difference between having a visual

---

[18] Harman, "The Intrinsic Quality of Experience," 39.

* According to Block (1996, 2003), Harman has confessed, in conversation, to holding a view similar to Lycan (1996).

** For example, Macphearson (2006) and Nickel (2006).

‡ For a survey, see Tye (2000).

experience of something versus an auditory experience of the same thing.[19] A simple version of the argument is as follows:

(1) It is obvious that our experience of touching a dog's fur and seeing the dog's fur are phenomenally different.
(2) However, we are representing just the same thing (the dog's fur) in both cases.
(3) Therefore, there is a phenomenal difference without a representational difference and Phenomenism wins.

This is a watered down version of the argument; however, what I want to highlight is that a *t-representationalist's* rebuttal is markedly different than a *functionalist's.** The representationalist argues simply that we represent different properties of the world across different modalities.[20] For instance, when we are looking at a dog's fur, we are representing its color, but when we are feeling the fur with our hands, we are representing its texture. Importantly, the types of properties we represent will never perfectly overlap (i.e., we cannot represent things like color in haptic sensations). Perhaps we can represent textures visually (for instance, having a visual experience of popcorn ceilings) but it is a far less detailed representation than the one arrived at through tactile stimulation. Given all the different kinds of properties available to our different sensory modalities, it seems likely (at least, it is arguable according to the representationalist) that there will never be the same representational content across modalities.

Block's challenge attempted to give a cross-modal example that reduced the amount of content. He gave the scenario of very quickly seeing something in your peripheral vision (say, off to your right, and absent of color, shape, and size properties), and having an auditory experience of the same scenario. Block presumes that the only available intentional content is something like "there is something *over there*" without any other properties as fodder for representing. Tye rebuts that there would at least be "loudness" being represented in the auditory experience.[21] At this point Block invites us to imagine abstracting away from details to merely something with the content: *as that location*. The response that *t-representationalists* have to give is that such abstraction is impossible. And, if we theoretically could abstract away to content like *the shape itself* (or something like that) it is not at all obvious that a "visual" or "auditory" experience would be phenomenally different.

---

[19] Block, "Mental Paint," 165-200.

* Of course, the functionalist can also make the t-representationalist's kind of rebuttal.

[20] Tye, *Consciousness, Color, and Content*, 45-54.

[21] Ibid, 45-54.

Functionalists respond to the challenge of cross modality by granting the possibility of identical representational content. What explains the phenomenal difference is found in the facts about how our visual system works, as opposed to how our auditory system works. Our visual representations seem so ***visual*** because, functionally, they are being operated by a system that functions very differently than an auditory system. The functionalist, then, has two different theoretical stances at his/her disposal:

> (1) Within a modality, phenomenal differences supervene on representational content. As in, for any visual experience (or auditory experience, or tactile experience, etc.), if the representational content is identical, the phenomenal content is identical. Or,
> (2) Within a modality, phenomenal differences supervene on representational and functional content.

The t-representationalist adopts (1), but rather inconsequentially. They are already committed to the claim that supervenience holds across modalities (so, *of course* they are committed to 1). Whether the functionalist adopts (1) or (2) is related to an ambiguity so far unresolved. This relates to the distinction Lycan makes of "qualia" versus "what it is like" or "overall feel." Adopting (1), for the functionalist, means that qualia are purely representational (i.e., they supervene only on intentional content) but that our "overall feel" supervenes on intentional and functional content. However, adopting (2) would change the thesis to allow for "qualia" to be representational and functional.* The motivations for these different views will be discussed later.

BODILY SENSATIONS AND MENTAL OIL

While Block argued that the properties of mental experiences that are doing the representing are mental paint, he also argued that certain experiences – like bodily sensations – contain experiential properties that are not representing anything. Experiencing these properties is experiencing mental oil. It is difficult to understand what mental oil is abstractly, but Block says that bodily sensations (the examples of pain and orgasms are the most pervasive in the literature) contain mental oil. Additionally, mental oil is the stuff about an orgasmic experience (or pain experience) that we enjoy the most (or the least).

The problem that bodily sensations present for t-representationalists is

---

* This is inconsequential to the functionalist the same way that (1) was for the representationalist; the changes within a modality are no different than changes cross modally in their supervenience requirements.

that it is not obvious what, if anything, our pain sensation is representing. The case for orgasms is similarly difficult in that it is not clear (a) what we are representing, and (b) how representing those things leads to the kind of experiencing we are having (these "phenomenally impressive" experiences). T-representationalists can posit intentional content like "something pleasing is happening *down there* that is changing in pleasingness – alternately increasing and decreasing in pleasure." The challenge Block puts forth is an intuitive one. Why should we think that the difference between me watching someone else orgasm (or thinking about orgasms abstractly), and myself having an orgasm is merely that the "there" is, well, *there*. T-representational answers can include things like "warmth, squeezing, throbbing, pumping and voiding."[22] Functionalist responses will turn to things like the functional differences between ascribing orgasms to someone else versus ascribing them to oneself as a part of what makes the experience so "impressive." Though there are disagreements, everyone seems to agree that, in the case of bodily experiences, it is not the case that you are experiencing a part of your experience (that is doing the representing). That is, they are not disagreeing about transparency.*

The different functionalist/t-representationalist rebuttals come out more clearly in pain cases. Similar to the orgasm cases, it is not obvious that our experiences of pain *represent* anything. T-representationalists are going to deny that "pain" is a kind of mental object that gets added to.[23] Instead, they will argue, "pain" is a token feeling, that is, at time *t*, it is one and the same as your experience at time *t*. Each token experience will represent different kinds of tissue damage (or so):

> a twinge of pain represents a mild, brief case of damage. A throbbing pain represents a rapidly pulsing disorder. Aches represent regions of damage inside the body rather than on the surface... A stabbing pain is one that represents sudden damage over a particular well-defined bodily region... In the case of a pricking pain, the relevant damage is represented as having a sudden beginning and ending on the surface or just below, and as covering a very tiny area. A wracking pain is one that represents that the damage involves the stretching of internal body parts (e.g., muscles).[24]

---

[22] Lycan, *Consciousness and Experience*, 136.

* See Tye, (2000); Block, (2003); and Lycan, (1998). They are all agreeing that such experiences are, more or less, explained in terms of transparency and without appeal to introspection.

[23] Tye, *Consciousness, Color, and Content*, 45-54.

[24] Michael Tye, "A Representational Theory of Pains and their Phenomenal Character" in *The Nature of Consciousness: Philosophical Debates*, ed. N. Block, O. Flanagan, and G. Güzelder (Cambridge, MA: MIT Press, 1997), 333.

The functionalist gives a different kind of answer:

> the (first-order) pain itself is constituted in part by its quale, which is its representational content, and in part by its functional role. A pain's giving rise to an introspective feeling and/or belief that one has the pain is but one among its other functional effects, like groaning, withdrawing, favoring, rubbing, the desire that it stop, the distraction and the rest.[25]

The functionalist then has the duty of explaining how functionalism and t-representationalism work in tandem and how the two could ever be dissociated in a principled way. This kind of response removes a part of the ambiguity we introduced above. If the functionalist is committed to this answer regarding bodily sensations, then it seems they are similarly committed to claim (2) above. Any phenomenal experience, like looking at a red ball, will be influenced by its quale and functional content. Because this is first-order, experiences of the kind "what it is like" are *further* influenced by other functional considerations. The supposed danger of such a view is that it seems to just collapse into functionalism. However, if your primary motivation for t-representationalism is to solve a metaphysical worry (as discussed above) then once the qualia are located, intentional content no longer has to do any heavy lifting.

FUNCTIONALISM

Something that has come out so far in our discussion is the remarkable similarity between the views held by Lycan and Block. Aside from the obvious stipulation that Block thinks there are characteristics of phenomenal experience that outrun even functional considerations, it is not clear how their theories are actually in conflict otherwise. This stipulation is found more in the inverted spectra arguments that have been neglected in this paper, and if we look at the arguments for mental paint and mental oil above (which have not mentioned outstripping functionalism), the difference looks largely *terminological*.

The source of contention between Lycan and Block concerns their views about functionalist theories of mind. Lycan is a card-carrying functionalist while Block is not. The debate about functionalism has a large and vibrant history that I cannot hope to do justice to here. However, a few words will help illuminate why their differences are not terminological but stem from deep philosophical concerns.

---

[25] Lycan, "In Defense of the Representational Theory of Qualia," 484.

The most famous worry pushed by Block is a kind of generality worry about functionalism.[26] Block argues that functionalism is stuck between a "pernicious liberalism" and an "unjustified chauvinism." If you set the relevant parameters for defining the functionality of a system too broadly or abstractly – i.e., an abstract relation between under-described inputs and outputs – then a whole host of phenomena can be attributed properties of mental states (like consciousness). Block's example is taking 1 billion (and change) inhabitants of China and outfitting them with two-way radios (each connecting one member to the relevant others). This whole system would eventually be connected to the relevant output nerves that control your body via radios. Block imagines that the inputs would be available via satellites that are theoretically viewable from anywhere in China.

Given the plausible realizability of this thought experiment, the functionalist is committed (if they define inputs and outputs too abstractly) to saying that China (or its participating members, anyway) have mental properties and, especially for our purposes, "qualia." This is, according to Block, *prima facie* unacceptable. In fact, if we define functionalism in terms of inputs and outputs, then there is no reason to think that certain economic systems (that follow the right pattern of monetary input and output) would not also fit a functional description identical to that of a mental one! However, if we attempt to restrict functionalism to species-specific properties (i.e., species that have *brains*), functionalism loses its dialectical role against physicalist accounts of consciousness. Therefore, functionalism is either too liberal (which is completely unintuitive) or too chauvinistic (and therefore indistinguishable from pure physicalist theories).

For the broader purposes of our discussion, the reason Block posits ontologically "new" properties of phenomenal experience (besides the inverted spectra objections which are arguably handled by the right kind of externalist theories about intentional content) is an *antecedent* rejection of functionalist accounts of the mind. In light of these considerations, Lycan and Block agree that merely intentional content cannot determine (or constitute) the nature of our experiences, they simply disagree on what things contribute the final phenomenal experience. They both fail to provide stories about how this extra (functional or phenomenal) content actually realizes in experience.

ATTENTION AND MENTAL PAINT

Our exposition now allows us to appreciate where Block's essay "Attention and Mental Paint" fits into the general dialectic between representationalist, functionalist, and phenomenist theories of experience. This is especially

[26] Ned Block, "Troubles with Functionalism," in *Perception and Cognition: Minnesota Studies in the Philosophy of Science, Vol. IX*, ed. W. Savage (Minneapolis: University of Minnesota Press, 1987), 261-63.

important because, I argue, Block takes his paper to apply to *strong-representationalist* views more broadly when it in fact does not. This half of my paper is devoted to applying Block's arguments in light of our foregoing discussion and then presenting a defense (or a direction toward a plausible defense) for representationalist theories.*

Block's 2010 argument takes a different approach than the type of examples we've seen so far. As opposed to arguing that there are parts of our experience that "outrun" intentional and functional content, he argues that our phenomenal experience is "incompatible" with representational content. It is incompatible according to the following dilemma: either,

> (1) Representationalism cannot account for changes in phenomenal experience by appeal to intentional content (so, representationalism is false), or,
> (2) The only strategy that it has, according to Block, to account for the phenomenal change, makes it incompatible with the phenomenal character of our experience more broadly.

The force of (2) is that should the representationalist adopt the strategy of "vague" contents to account for the phenomena Block introduces, it has the problem that our experience is not "vague." However, Block admits that his dilemma relies on several crucial premises:

> (A) The transparency of experience.**
> (B) That representationalists cannot rely on veridicality conditions (i.e., illusion or misrepresentation) to explain the phenomena.

His first premise (A) is a bit confusing. Block is arguing for mental paint, which has traditionally been a rejection of the negative thesis of transparency (that we cannot be aware of the experiences themselves). Mental paint properties were properties of the experience that had been doing the representing that were affecting our phenomenal experience. Here the claim is the following: the transparency thesis holds that our experience is the experience of the properties of external objects (that we are representing). However, the apparent properties of those external objects change (I have not shown how Block argues this, yet);

---

* Block also attacks Direct Realism; however, insofar as representationalism is a more common (and defensible) theory, I will be talking about representationalism instead.

** This is likely only a premise because it is what his opponents hold; however, it is worth examining its relation to mental paint and mental oil arguments.

therefore, something other than intentional content must be playing a role in our supposed direct (or diaphanous) experience. Though the claim seems clear enough, it is difficult to see how it is not actually the rejection of the negative transparency claim unless it is something *like* mental oil. That is, transparency must be something that Block is arguing against on the basis of contradiction (thereby proving mental paint or an equivalent), or he must be abandoning the claim of mental paint for the claim of mental oil (which he does not purport to do).

Block attempts to avoid this by splitting the transparency thesis into two claims, a positive and negative one. The positive claim is the diaphanousness of experience (that our qualia are of external objects) while the negative claim is the "strong" view I attributed to Tye and Dretske – that qualia are the only aspect of experience. He claims that the positive view alone will "hoist" representationalists by their own petard, but this is not the case. As long as you reject the 'negative' or 'strong' claim, Block's worries do not affect you. The positive claim is unfit as a premise because it is uncontroversial and, as has already been said, cannot do the work in the argument alone.

The importance of this exercise is to understand who Block's target is. Block takes himself to be targeting a whole host of opponents, including Tye, Dretske, and Lycan. However, the differences illuminated earlier in this paper show that, if his first premise is actually a rejection of the negative claim of transparency, or a claim about mental oil, then he is, in fact, not arguing against functionalist theories! It is difficult to see what he could be arguing unless it is one of the claims sketched above, but later in his paper, he says he is arguing against "a version [of representationalism] in which phenomenal character is determined by or flows from content."[27] This is clearly not a functionalist view. Therefore, even if Block's arguments go through successfully, he does not argue for any type of ontologically "new" part of experience.*

The question now worth exploring is how threatening this argument is for strong transparentists (or t-Representationalists) like Tye and Dretske. Block's entire argument hangs on premise B. This is different than past papers which have had little or nothing to do with the notion of veridicality. Before going into potential representational responses, we should look at his argument.

ATTENTION

Block brings a lot of recent evidence from experimental psychology about attention to bear on the debate. Specifically, Block provides evidence for the following:

---

[27] Block, "Attention and Mental Paint," 27.

* Without the antecedent arguments against functionalism already discussed.

(1) We can consciously see things we are not fixating our eyes on (like looking at something out of the corner of your eye).
(2) One can see something without focally attending to it as well as without fixating on it.
(3) We can focally attend to something without fixating on it.
(4) Focal attention changes perceptual qualities such as: "perceived contrast, perceived color saturations, perceived object size, perceived spatial frequency (stripe density), perceived gap size, perceived speed, and perceived flicker rate."[28]
(5) Attentional effects, both excitatory and suppressive, pervade the visual field.

The first three of these are not worth going into much detail about here, and I will take them as given for the purposes of our discussion.* The importance of the first three amounts to restricting a certain type of representational retort to many counter examples. That is, changing your eye direction allows new and different information to pervade your visual field. If there is a phenomenal difference as a result, it is easy to point to new information (and therefore, new intentional content) as the constitutive reason.

The fourth empirical claim has been well established by a cadre of experimental psychologists. Though Block gives numerous examples, we can focus on one of the more important ones. The figure below (Figure 1) is a figure of two "Gabor" patches – one at 22% contrast and the other at 28% contrast (the 6% contrast difference is easy to see). The effect reported by Carrasco, Ling, & Read is one in which involuntary attention attracted to the left increases the apparent contrast of the 22% patch (and decreases the apparent contrast of the 28% patch) to the point of subjective equality between the two patches.[29] With a bit of practice (remember claim 3), one can voluntarily shift attention to the left or to the right and experience the increase in apparent contrast from this diagram. The square dot in the middle is the fixation point. If you were to attend to the 28% patch, it would look even higher in contrast, and the 22% patch would look even lower in contrast. This effect has been seen for a number of apparent properties listed in (4).[30]

---

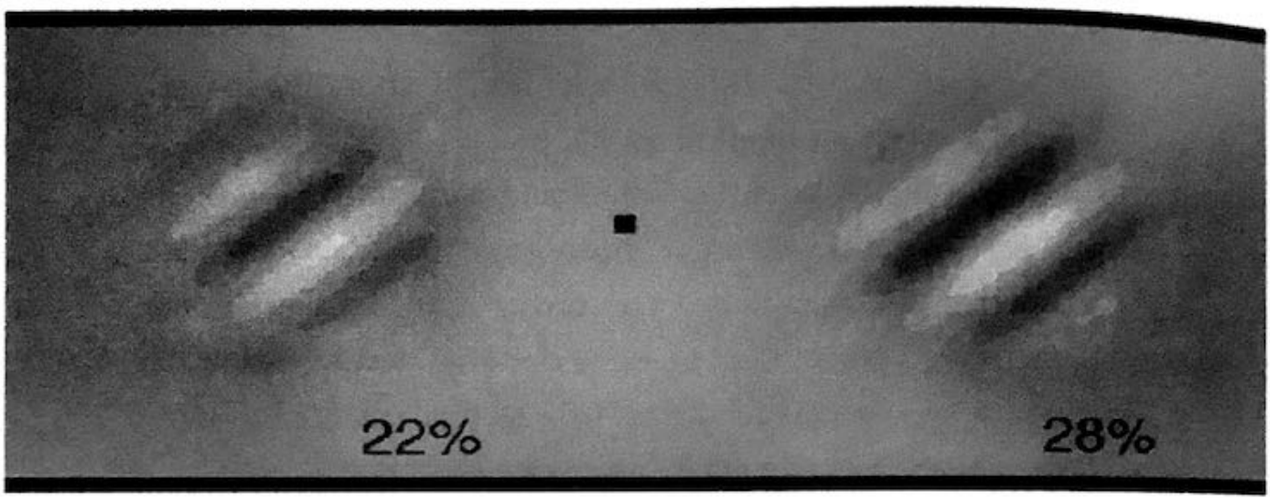
[28] Block, "Attention and Mental Paint," 33.

* Michael Tye (2009b, 2010a) argues that being able to attend to something is necessary for (consciously) seeing it. This is a debate above and beyond the scope of this paper.

[29] M. Carrasco, S. Ling, and S. Read, "Attention alters appearance," *Nature Neuroscience* 7(2004): 312.

[30] J. Gobell and M. Carrasco, "Attention alters the appearance of spatial frequency and gap effect," *Psychological Science* 16(2005): 644–651.

FIGURE 1



Block also goes to lengths to assure us that this effect is a *perceptual*, rather than cognitive effect. Although there is some evidence for this in the fact that the effect only worked involuntarily if the stimulus was presented in under 53 milliseconds (and therefore far too short a time for cognitive processes to do any heavy lifting), the main source of evidence is examining attention as primarily a neural operation. For instance, attention to a stimulus boosts the firing rate of neurons that respond to that spatial area and suppresses the firing rate of neurons to other areas. Additionally, the extent to which these neurons are excited or suppressed reflects the degree of attention one is showing.[31] Attention is also subject to "adaptation." Adaptation is an effect on perceptual systems when, after prolonged exposure to a stimulus (this can be about a minute or so), the receptor response to a number of perceptual parameters (orientation, contrast, etc) decreases with respect to those parameters.[32] This can be easily demonstrated by looking at some of the vertical grating examples in Block's paper. This is also the effect of looking at the night sky. If you fixate on a single star for about a minute you no longer see it (however, if you move your eyes and look with your peripheral vision, for instance, the star reappears).

The last facts about attention Block brings to the table are summarized in (5). That is, current researchers on attention have rejected the "steady" spotlight model[33] and its slightly younger cousin, the "gradient of attention" model.[34] The problem with these models is that they assumed a single "field" of attention that would expand or contract in conjunction with attending more narrowly (with more detail) versus more broadly (with less detail but more objects). Instead,

---

[31] J.H. Reynolds, "Mapping the microcircuitry of attention," *Nature Neuroscience* 11.9(2008): 861-62.

[32] P. Series, A.A. Stocker, and E. Simoncelli, "Is the Homunculus 'Aware' of Sensory Adaptation?," *Neural Computation* 21(2009): 3271-72.

[33] M. Posner, C. Snyder, and B. Davidson, "Attention and the detection of signals," *Journal* of Experimental Psychology: General 109(1980): 172-74.

[34] C.J. Downing and S. Pinker, "The spatial structure of visual attention" in *Attention and Performance XI*, ed. M. Posner & O.S.M. Marin (London: Erlbaum, 1985), 183–87.

attention has two independent fields: an excitatory field and a suppressive field. Secondly, each field is quite large and irregular.* Finally, attentional resources are shared with other aspects of perception, like attention in visual processes and auditory processes, executive control mechanisms,** and cognition. This last part is not surprising to anyone familiar with inattentional blindness (or simply talking on a cell phone).

ILLUSION

In response to these facts the t-representationalist has a seductively easy recourse: the representationalist can simply say that the apparent increases in contrast, size, speed, etc., are all simply due to changes in our representational content – after all, that is what representationalism (even broadly construed) is! They would claim that different distributions of attention yield different representational contents. The problem with this proposal, according to Block, is that this strategy only works if it is grounded in a concept of veridicality. Here is Block's challenge:

(1) T-representationalist is changing intentional contents at will to fit appearances (this is the representationalist thesis).
(2) If the layout is the same then the representationalist cannot pick different properties to represent.
(3) If (2), then at least some of the representations (i.e., one of the percepts – in our case, the 22% patch or the 28% patch) must be illusory.
(4) (2) - we are in fact fixating on the same location and merely shifting attention.
(5) But, neither precept is illusory.
(6) Therefore, t-representationalism cannot account for the differences (and is false).

The work to be done in this proof is with regard to premise 5 (this corresponds to premise 2 of the essential premises mentioned earlier). The first four are *prima facie* reasonable and will remain unquestioned for now. Block's argument for (5) relies on the problem of picking a "cut-off" that properly engenders veridical perception and illusion, respectively.

Block examines the simple view that what defines veridical perception is an absolute value of attention on an object. This is clearly a problem, because

* See Datta & DeYoe (2009); Hopf et al., (2006).
** See Brand-D'Abrescia and Lavie (2008).

attentional resources are shared with many other faculties, and would therefore entail that were I talking on a cell phone, none of my visual experiences would be veridical. Therefore, Block supposes, we should switch to a system based on "relative allocations of attention." But, says Block, the points about the new model of attention show that there are many areas of our visual field that are very close in the amount of attention allocated to them compared to our supposed focus of attention. Making such small differences (and in everyday circumstances), the difference between veridicality and illusion makes no sense. In order to say that one patch (the 22% or the 28% patch) is an illusion requires something like one of these non-context-relative versions of veridicality that, as we can see, suffer unintuitive consequences. The easier path, for Block, is just to say that veridicality does not apply to cases of mere attentional allocation because of mental paint (that is, they can be context relative). Unfortunately, the representationalist needs illusion in order to have two contradictory properties applied to the same external object.

Block has highlighted an important idea about our visual cognition: things will, across a confluence of obvious and hidden variables (like attention but I imagine it applies to a lot more), look slightly different (even if we do not judge them to). This fact, all by itself, makes a theory of veridicality difficult for the representationalist.* There are, however, a couple of considerations with which the representationalist can respond.

First off, the representationalist can deny that we are looking for a theory that will allow us to say that one precept is illusory while the other is not. They can instead focus on whether or not specific properties (just those properties being represented) are illusory. For a simple example, let's say we are playing catch and represent the ball as a red ball (when it is supposed to be white). In one sense it is fair to say that I had suffered from an illusionary precept of a red ball; after all, what I saw was not true. However, it is not totally fair to say that my whole precept was illusory because it was in fact a *ball*. And, I got the relative location correct (I caught it), in addition to a whole host of other properties about the ball that I correctly perceived.

Given that our interest is in the properties of objects (not some holistic veridicality), of course a representationalist would not define veridicality in terms of levels of attentional allocation, but in terms of whether or not the properties were being represented accurately. The representationalist is more than free to say that certain properties do not require full attentional resources to be accurately represented. For instance, I can see a round thing in my

* I say representationalist here, as opposed to t-representationalist, because this more general concern about veridicality applies to functionalist views as well. For the rest of the paper I will be referring to representationalist as more broadly defined.

peripheral vision and accurately represent that it is round. This may mean that a lot of the things we represent (i.e., experience) are not veridical, but they are usually not important or do not impede our functioning. While they come out in super-specific situations like contrasting Gabor patches, we usually are working with a lot more properties. This would be some sort of error theory that is not much more threatening than the "grand illusion" hypothesis.[35] The blow may be lessened, I imagine, by saying that our visual processing probably gets most things veridically when it is allowed to function normally. That is, we are not fixating on the same spot for a long time, and our eyes are allowed to gather much more information about the scene.

These considerations allow the representationalist to say that that they have a theory of context-free veridicality, which claims some properties represented in our phenomenal experience are illusory, but they cannot tell you which ones (and they do not have to).* They have rejected (5) and are not threatened by Block's proposal.

The representationalist may also push back on premise (4). This premise states that the "layout" of our visual scene is, in fact, the same. The representationalist claims that the fact of our neuron's responding faster or slower given certain attentional foci results in a change of information or "layout." That way, the representationalist can have a contextual veridicality (like the one Block wants), and say that all the precepts are veridical relative to the information they are receiving. This seems like a long shot so I will not attempt to argue for it here.

LOUDNESS

One may wonder why Block claimed there cannot be a context-free kind of veridicality in these cases. After all, it causes Block to admit that it leads to the "paradoxical" conclusion that neither percept is illusory *and,* as we saw, the representationalist has an easy answer to the challenge. The answer, if there is one, is to be found in Block's analogy of "loudness." Loudness – the perceived intensity of a sound – is a quality that presents intensity but is also the product of frequency, bandwidth, and duration. He claims:

> Although loudness in some sense represents intensity and is experienced as presenting intensity, the same intensity can sound differentially loud depending on other variables. Analogously, although perceived size presents actual size, [it]

---

[35] A. Noe, "Is the visual world a grand illusion?," *Journal of Consciousness Studies,* 5.6(2002): 1-2.

* So, it turns out, representationalists can postulate intentional contents at will.

> is a function not only of the actual size but of other variables, notably the distribution of attention.[36]

Block wants to argue that "perceived size" (or contrast, speed, etc.) may bear the same relation to actual size as loudness bears to actual intensity, but notes that there is an important disanalogy between loudness and perceived size:

> the auditory system is built to register loudness and that is revealed in the fact that there is loudness constancy. A sound maintains its loudness as you move closer to the source and as its intensity increases. But the phenomena I have been describing involving attention have a kind of *anti-constancy* in that the perceived size changes even when the actual size is constant.[37]

The idea is still somewhat mysterious, but relies on the idea from Burge that perceptual objectivity is grounded in "constancy." Constancy is the capacity "systematically to represent a given particular, property, relation, or kind as the same, despite significant variations in registration of proximal stimulation."[38] For example, people show constancy by representing a certain color as the same despite changes in light intensity; or, representing something as being the same size despite it taking up different amounts of our visual field.

The idea pushed by Burge and Block, if I understand them correctly, is that it does not make any sense to apply notions of veridicality to perceptual states that do not show constancy. That is, veridicality only makes sense when applied to states that purport to be objective. A necessary condition for objectivity is constancy. From Burge this stems from a much larger project that involves finding the origins of object perception and representation in comparative ethology and evolution. His project jettisons many of the ways that philosophers (including the representationalists in this paper) have thought about representation. Without going into his project, we can see the problem manifest in the previous section, where I mentioned how Block highlighted the worry that our phenomenal experiences undergo constant change.

Unfortunately, the representationalist, given the kind of response sketched in the previous section, has little reason to opt for a constancy requirement. Block, instead of arguing that that kind of view is mistaken, pushes the buck to Burge whose project bears little resemblance to the historical debate sketched in

---

[36] Block, "Attention and Mental Paint," 25.

[37] Ibid., 55.

[38] T. Burge, "Perceptual Objectivity," *Philosophical Review*, 118.3(2009): 319.

the previous sections of this paper.* Also, and more to the point, Block has not actually explained how an experience can represent X as being F, and another experience represent (the unchanged) X as something incompatible with F, while neither experience is illusory.

CONCLUSION

Block's paper "Attention and Mental Paint," in the context of the larger debate between representationalists and phenomenists, provides an extremely narrow argument. His argument targets *only* representationalists, and only on the grounds that they cannot provide adequate conditions of veridicality. When we examine the troubles for veridicality, we find that the representationalist has an easy answer that grounds veridicality in the properties that objects actually have – as opposed to in varying levels of attention. Any motivation for abandoning this kind of veridicality (which is applied even in the face of ever-changing and fickle phenomenal experiences) stems from anti-individualist considerations presented in Burge, but not argued for by Block himself. The kinds of considerations that may sway those committed to the traditional approaches in the philosophy of mind (i.e., the kinds of arguments and positions given throughout this paper) are many, complicated, and beyond the scope of my analysis, but must be examined before Block's newest argument for mental paint can be appreciated.

* It should be noted that Burge is a 'weak-representationalist' like Block.

BIBLIOGRAPHY

Block, N. "Troubles with Functionalism." In *Perception and Cognition: Minnesota Studies in the Philosophy of Science, Vol. IX*. Edited by W. Savage. Minneapolis: University of Minnesota Press, 1987. 261-325.

Block, N. "Inverted Earth." *Philosophical Perspectives* 4(1990): 53-79.

Block, N. "Mental Paint and Mental Latex." *Philosophical Issues* 7(1996): 19–49.

Block, N. "Mental Paint." In *Reflections and Replies: Essays on the Philosophy of Tyler Burge*. Edited by M. Hahn and B. Ramberg. Cambridge MA: MIT Press, 2003. 165–200.

Block, N. "Attention and Mental Paint." *Philosophical Issues* 20(2010): 23–63.

Brand-D'Abrescia, M., and N. Lavie. "Task Coordination between and within sensory modalities: Effects on distraction." *Perception & Psychophysics* 70.3(2008): 508–515.

Burge, T. "Perceptual Objectivity." *Philosophical Review* 118.3(2009): 285–324.

Carrasco, M., S. Ling, and S. Read. "Attention alters appearance." *Nature Neuroscience* 7(2004): 308-313.

Carruthers, P. *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: Cambridge University Press, 2000.

Datta, R., and E. DeYoe. "I know where you are secretly attending! The topography of human visual attention revealed with fMRI." *Vision Research* 49(2009): 1037–1044.

Downing, C. J., and S. Pinker. "The spatial structure of visual attention." In *Attention and Performance XI*. Edited by M. Posner & O. S. M. Marin. London: Erlbaum, 1985. 171-187.

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

Gobell, J., and M. Carrasco. "Attention alters the appearance of spatial frequency and gap effect." *Psychological Science* 16: (2005) 644–651.

Harman, G. "The Intrinsic Quality of Experience." *Philosophical Perspectives* 4 (1990): 31–52.

Harman, G. *Reasoning, Meaning, and Mind*. Oxford: Oxford University Press, 1999.

Hopf, J.-M., C.N. Boehler, S.J. Luck, J.K. Tsotsos, H. J. Heinze, and M.A. Schoenfeld. "Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision." *Proceedings of the National Academy of Sciences* 103.4(2006): 1053–1058.

Lycan, W. G. *Consciousness and Experience.* Cambridge, MA: MIT Press, 1996.

Lycan, W. G. "In Defense of the Representational Theory of Qualia (Replies to Neander, Rey, and Tye)." *Nous* 32.s12(1998): 479-487.

Lycan, W. G. "Representational Theories of Consciousness," *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition).* Edited by Edward N. Zalta. <http://plato.stanford.edu/archives/fall2008/entries/consciousness-representational/>.

Macpherson, F. "Ambiguous Figures and the Content of Experience." *Nous* 40.1 (2006): 82–117.

Nickel, B. "Against Intentionalism." *Philosophical* Studies 136(2006): 279-304.

Noe, A. "Is the visual world a grand illusion?" *Journal of Consciousness Studies* 5.6(2002): 1-12.

Posner, M., C. Snyder, and B. Davidson. "Attention and the detection of signals." *Journal of Experimental Psychology: General* 109(1980): 160–174.

Reynolds, J. H. "Mapping the microcircuitry of attention." *Nature Neuroscience* 11.9(2008): 861-862.

Series, P., A.A. Stocker, and E. Simoncelli. "Is the Homunculus "Aware" of Sensory Adaptation?" *Neural Computation* 21(2009): 3271–3304.

Tye, M. "A Representational Theory of Pains and their Phenomenal Character." In *The Nature of Consciousness: Philosophical Debates.* Edited by N. Block, O. Flanagan and G. Güzelder. Cambridge, MA: MIT Press, Cambridge, Massachusetts: MIT Press, 1997. 329-340.

Tye, M. *Consciousness, Color, and Content.* Cambridge, MA: MIT Press, 2000.