Cosmos and History: The Journal of Natural and Social Philosophy, vol. 16, no. 2, 2020

MENTALITY AND OBJECT: COMPUTATIONAL AND COGNITIVE DIACHRONIC EMERGENCE

Ekin Erkan

ABSTRACT: Espousing non-reductive physicalism, how do we pick out the specific relevant physical notion(s) from physical facts, specifically in relation to phenomenal experience? Beginning with a historical review of Gilbert Ryle's behaviorism and moving through Hilary Putnam's machine-state functionalism and Wilfrid Sellars' inferential framework, up to more contemporaneous computationalist- and cognitivist-functionalism (Gualtiero Piccinini), we survey accounts of mentality that countenance the emergence of mental states vide input- and output-scheme. Ultimately arriving at the conclusion that functionalism cannot account for problems such as no-cognition reports, we see any robust defense of physicalism must appeal to other principles. Thus we move on to the question of emergence, not as it pertains to the hard(er) problem, but to the matter of conceptual externalization of mental properties from physical properties. Accordingly, we navigate Karen Bennett's compatibilist solution to the exclusion argument against mental causation for the non-reductive physicalist position, according to which the physical effects of mental cases are not overdetermined, demonstrating that this backfires by offering a path for the mind-body interactionist Dualist to claim causal closure by appealing to this same schema. We conclude with a series of conceptual musings regarding rationality which take into account our challenges and findings, querying about whether phenomenal consciousness is a fundamentally private, or socially configured, notion.

KEYWORDS: Philosophy of Mind; Epistemology; Machine-State functionalism; Kant; Wilfrid Sellars; Non-reductive Physicalism; Emergence

> The **immediate** consciousness of the existence of outer things i s not presupposed but proved in the preceding theorem, whether we have insight into the possibility of this consciousness or not. The question about the latter would be whether we have only an inner sense but no outer one, rather merely outer imagination. But it is clear that in order for us even to imagine something as external, i.e., to exhibit it to sense in intuition, we must already have an outer sense, and by this means immediately distinguish the

> mere receptivity of an outer intuition from the spontaneity that characterizes every imagining. For even merely to imagine an outer sense would itself annihilate the faculty of intuition, which is to be determined through the imagination outer intuition from the spontaneity that characterizes every imagining. For even merely to imagine an outer sense would itself annihilate the faculty of intuition, which is to be determined through the imagination.
>
> —Immanuel Kant, The Critique of Pure Reason, B276-7.

INTRODUCTION: EXCLUSION

Reza Negarestani's *Intelligence and Spirit*, multitudinous, variegated, ambitious, and a bit disheveled in its Hegelian-Carnapian synthesis, perhaps can be best summarized as attempting to provide a sundry methodology to consider how we might approach the rather awesome question: ought Intelligence, as it is collectively situated vide the Dasein of Geist—through linguistically licensed epistemic practices including the verification and articulation of knowledge and reasons—be considered as fundamentally private, subjective affair or is it inherently deprivatized? The latter position, which Negarestani espouses, implies that intelligence is structured not as any one property but a shared and distributed representation of a network of lower-level representations; the qualitative integration of self-consciousness or Geist transmogrifies mind itself into a cognitive artefact of intelligence. Semantically structured, the conclusion that Negarestani's book offers is that the function of the mind that is "intelligence" is a semantically structured and negotiable concept, stilted on a logical space of reasons and, therefore, but an ossified model for thinking collective agency. In short, intelligence is but a socially distributed metaphor, imposing a boundary-distinction between intelligence understood as the intractable organ of thought, and intelligibility—the concrete actualization of intelligence that renders the world and sapience thinkable.[1]

This is a kind of understanding as intelligence as structuring-activity: thus, intelligence finds a critical node with the first instantiation of tool-use, reified and articulated in distinct terms with every novel technology thereafter, revealing the manifest world by rendering its scientific counterpart intelligible, modeling reality on the normativity of the space of reasons. Accordingly, structuration is then identified with conceptualization and objectivation: making-intelligible a reality

[1] Reza Negarestani, *Intelligence and Spirit* (Falmouth: Urbanomic, 2018).

that is in itself neither functionally structured nor conceptually intelligible. At the same time, avowing the pertinence of AGI as the theoretical medium through which Geist is formalized, Negarestani reveals his Carnapian penchant. This proclivity is rather straightforward, as it deals with providing an external framework condition; privy to Carnap's project of descriptive semantics—the pragmatic use of propositions in making cognitive judgments in suitable perceptual or experimental contexts about localized individuals/particulars—within non-formal domains, Negarestani's Carnapian flank concedes that, when it comes to semantic externalism, verification empiricism can at best aspire to make cogent conceptual explication, where our conceptual explications, viz. *explicandae*, are in principle incomplete and thus revisable. Thus and so, our explications must clarify and improve upon their *explicatae* within their original contexts of use. Ineluctably, this invokes semantic and justificatory externalism. Tool use, semantics, artworks, ethics—when instantiated as particulars, these are all indexes of intelligence, if we adopt Negarestani's framework.

What of Hegel, who Negarestani often returns to? In the service of conceptualizing intelligence, Hegel's phenomenological method ineluctably involves positive conclusions, augmenting Kant's Critical account of rational judgement and justification. When capering alongside Hegel's spirit, absolute contingency becomes equipollent to absolute necessity with the totality of form and any one-world system including the totality of every possibility whatsoever. This all much to Quine's chagrin as Quine, wielding Russell's pickaxe—that theory of descriptions meant to untangle the matted knots of Plato's beard—warn us from inviting Hegel to dine; undoubtedly, Hegel's invitation also means welcoming unactualized modal possibilities. For Hegel, actuality includes its opposite, actuality met with the conditions of possibility, attempting to actualize possibility, itself; the modal ramifications of substance expound via both actuality and possibility, with that which is unactualized embedded as conditions of possibility. The form of consciousness Hegel critiques as 'sense certainty' espouses naïve realism, with Hegel's phenomenological presentation highlighting those varieties of mediation involved in what sense-data-cum-sense-certainty purports as utterly immediate, and thus fallacious, knowledge. For those mediations that are conceptual and cognitive involve competent use of various concepts in

connection with sensed particulars.[2] The inferential gambit of Negarestani's Hegel is to serve at once as a suspicion towards commonsense knowledge of anything—although we do know *how* to and do indeed engage in sorting out and sequencing our experiences of various particulars and experiential episodes, this is not simply done through sensing. Notably, however, Negarestani's uptake allots us something more akin to a Fregel,[3] that mongrel-philosopher who so seeks to identify accurately sub-personal cognitive functions which must be sufficiently reliable so as to enable us to be semantically aware of our epistemic warrants.

Let us, at this point, set Negarestani's project aside but glean from it one guiding methodological question: can intentionality be deprivatized? In non-formal domains, justification is not sufficient for provability and, vide Kant's Transcendental Analytic and the Analogies of Experience, we recognize that non-formal reasoning requires semantic and existence postulates, none of which can be formulated, assessed, or justified by the formal techniques of logic alone. This paper, which begins by clarifying a historical survey within the philosophy of mind before moving on to engaging with inferentially-licensed machine-state functionalism and non-reductive physicalism, is guided by the aforementioned set of conceptual concerns. However, given our limitations at this time, our considerations shall be significantly more modest than those aforementioned philosophers. Thus, we shall not even entertain intelligence, a notoriously slippery concept, and the possibility of deprivatized intelligence—instead, we shall turn to the question of conscious and mental events.

---

[2] G.W.F. Hegel, *The Phenomenology of Spirit*, trans. A.V. Miller (Oxford: Oxford University Press, 1978), §238.

[3] That Hegel unspooled by Brandom, named Fregel, queries whether the alethic modal dependencies through which we conceptually track regularities in the regular descriptive order can be set to enjoy real counterparts in nature? What of modality, logical negation-based differences—is modality only applicable to the domain of cognizing or is the world conceptually and modally-differentially structured? If the latter obtains, how do we conceive of this difference? Negarestani points to interaction/cut-elimination by way of Jean-Yves Girard's ludics. According to the inferential schema of ludics, contexts are cumulative in multiplicatives and shared in additives; negation (⊥) does not proffer the familiar role of "negating" but the role of a change of viewpoint. Negarestani's project thus leaves the reader baffled as to how the writer did not consider to reinvite and naturalize deontological metaethics into this formal scheme, perhaps offering a formalized developmental rejoinder to the Categorical Imperative and, thereby, answering Filipa Foot's charge re: the non-hypothetical use of "should," of how a self-given law that is not a physical law like the law of gravity, i.e., a law of behavior, can be universalized vis-à-vis pure rationality. Filippa Foot, "Morality as a System of Hypothetical Imperatives", *The Philosophical Review* 81, no. 3 (1972): 305-16.

COMMITMENTS AND DESTINATIONS

To approach this task we shall consider non-reductive supervenience physicalism. This meaning putting aside the metaphysical possibility of Davidson's token-physicalism and the consequent epiphenomenalism commitment, where mental processes are likened to smoke emanating from a locomotive. We grant that not only are mental processes significant insofar as phenomenological consciousness is concerned, but that they are causally efficacious, propelling the physical body forward. To lay bare our commitments, we presume that while phenomenal consciousness, or, per Ned Block's parlance, p-consciousness, is significant and that there is such a thing, it is compatible with the doctrine of physicalism. P-consciousness, we maintain, is separable from cognitive access. For instance, when observing a complex visual scene, we are phenomenally conscious of more than we can report. However, contra the "no-report" paradigm, which separates the neural basis of cognitive processes underlying post-perceptual decisions and reports from the neural basis of conscious perception, we, commit to the fact that, even in the absence of reportability, subjects may engage in some post-perceptual cognitive processing. Thus, we commit to a *no-cognition paradigm*, as evidenced by empirical cases such as binocular rivalry.[4]

Given no-cognition paradigms, we, more broadly, are committed to the possibility of totally inaccessible experiences, while also holding that neuroscience may, and most likely will, eventually support a theory on which the neural machinery that grounds (phenomenally) conscious experience is completely separable from the neural machinery that grounds cognitive access. As far as P-consciousness is concerned, there is something about our internal biological makeup, perhaps its electrochemical character, that is essential to conscious phenomenology. Thus, phenomenal experiences would, too, be separable from superficial functional organization, from externally determined

[4] "According to the prevailing account of rivalry, pools of neurons representing each of the incompatible stimuli inhibit one another. In the presence of neural noise, one pool wins temporarily. Then that pool is weakened by adaptation and the other pool representing the other alternative takes over. Because of the impact of neural noise, the time of the transitions cannot be predicted on the basis of past transitions. Binocular rivalry occurs in many animals, including fruit flies, and can occur in humans with invisible stimuli, showing that binocular rivalry is not intrinsically a conscious process." Ned Block, "What Is Wrong with the No-Report Paradigm and How to Fix It," *Trends in the cognitive sciences* 23, no. 12 (2019): 1004.

representational content, and from cognitive access, because they are constituted by internal biological states of the brain, and those biological states are separable from all those things. However, the possibility of this epistemic gap between those biological states in the brain and conscious experience does not commit us to an ontological gap.

SETTING UP THE PLAYING FIELD

Before dealing with the aforementioned conceptual matters, there is some need to Consider the possibility, as many philosophers have asked to consider, of a superficial functional isomorph of ourselves occupying another distant, possible world. This thought-experiment could be brought a bit closer to our actual reality, and thus our actual world, if, perhaps, we ask that we transpose this situation by way of teletransportation, as Derek Parfit asks to imagine in his 1984 classic *Reasons and Persons*. While we do not yet possess such a technology, it is not terribly unfeasible to consider the possibility of such a development in the not-too distant future. In Chapter 10 of *Reasons and Persons*, "What do We Believe Ourselves to Be?", Parfit asks us to imagine our teletransportation to Mars—in this world, teletransportation is possible and the way in which it is done is that, when we wants to travel to Mars we step inside of a teletransportation device and press a green button. Accordingly, our body and brain are destroyed in the flash of an instant, we lose consciousness, and the exact state of all of our cells is recorded. This information is, simultaneously, sent via radio to Mars, and on Mars, a machine utilizes this information to create a brand new body and brain out of new, though apparently identical, matter. Hence, we, in our new albeit identical body on Mars, consider ourselves to be the identical person when we wake up: our memories are retained, we behave the same way, have the same tastes, desires, proclivities, etc. Because memories are retained, there is psychological continuity between me and my replica; because the physical make-up is retained, there is physical continuity as well.

Now, we are asked to imagine the same scenario where we do not go unconscious during the teletransportation process but the body and that our body stays on Earth while the same process of replication occurs; this time, however, it is a process of duplication. We, observing via camera, see this replica of ourselves on Mars, going about their day. Parfit notes that, most likely, we would, in this second scenario not consider this person to be an identity. The stakes could be

raised ever higher—consider that the operator of the teletransportor tells you that something has gone awfully wrong and you suffered terminable heart damage, although your replica is fully healthy and did not. Parfit remarks that:

> Since I can talk to my Replica, it seems clear that he is not me. Though he is exactly like me, he is one person, and I am another. When I pinch myself, he feels nothing. When I have my heart attack, he will again feel nothing. And when I am dead he will live for another forty years. If we believe that my Replica is not me, it is natural to assume that my prospect […] is almost as bad as ordinary death. I shall deny this assumption. As I shall argue later, I ought to regard having a Replica as being about as good as ordinary survival. I can best defend this claim, and the view that supports it, after briefly discussing part of the past debate about personal identity.[5]

Thus opens the philosophical problem of a reductionist vs. non-reductionist view of personal identity. If we espouse the non-reductionist position, we are not satisfied with continuity by way of reducing identity to physical or psychological continuity—brains, bodies, and mental/physical events. For the non-reductionist, despite physical replication, some further fact needs to account for personal identity. According to the non-reductionist's view, personal identity over time does not simply consist in physical and/or psychological continuity and, thus, the non-reductionist searches for some further fact—a further fact which states that the criterion for personal identity is a separate entity or property that exists apart from the body and brain. Could this further fact incorrigibly appeal to something like "personal identity"?

In *The Conscious Mind* (1996), David Chalmers remarks that "[t]here is a strong intuition that there must always be a fact of the matter about personal identity: if there are numerous minds descending from my current state, there must be a fact about which one of them will be me."[6] This idea is precisely what Parfit criticizes, arguing persuasively that there is nothing more to the fact of personal identity than facts such as psychological continuity, memory, and the like. If we accept Parfit's analysis, then each of "tomorrow's minds," so to speak, are equal candidates to count as me, and there is no fact to distinguish them. In turn, the determinate "flow" of personal identity is but an illusion.

---

[5] Derek Parfit, *Reasons and Persons* (Cambridge, MA: MIT Press, 1984), 180.

[6] David J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory* (Oxford, UK: Oxford University Press, 1996), 352.

RYLE: ANALYTICAL BEHAVIORISM

Now, say that in this process of teletransportation, my replica is reconstructed not in Mars but in a planet that is analogous to Earth, perhaps $Earth_2$. Consider, furthermore, that on $Earth_2$, my reconstructed self has a different chemical basis—perhaps silicon instead of organic carbon. This is a thought experiment that Ned Block asks us imagine, considering Commander Data from Star Trek:

> Let us think of Commander Data as defined as a merely superficial functional isomorph of us. A superficial isomorph of us is isomorphic to us in causal relations among [its] states, inputs, and outputs to the extent that those causal relations are part of commonsense psychology. That is, for every human mental state, input, and output, there is a corresponding state, input, and output of Commander Data; and for every causal relation among our states, inputs, and outputs, there is a corresponding causal relation among Commander Data's mental states, inputs, and outputs. One consequence is that Commander Data will behave just as we do, as far as we can tell from the standpoint of commonsense psychology. I said that Commander Data is a merely superficial isomorph of us. That means that he is not like us in physical realization of the superficial functional states. … We might suppose just to get an example on the table that the physical basis of Commander Data's brain is to be found in etched silicon chips rather than the organic carbon basis of our brains … We have no conception of a ground of rational belief that Commander Data is or is not conscious.[7]

Commander Data and the $Earth_2$ Replica are superficial functional isomorphs. Functionalism, in both cases, allows for us to contend with the Multiple Realizability of mental states. Consider that Commander Data and the Replica can both feel pain when reacting to bodily damage; how might we know this? We see that, perhaps, both of them wince when they bump their head against a sharp object, or perhaps they both curse aloud, reporting that they are in pain while doing so. Given the network of pain-behavior, this scheme localizes pain as the center fulcrum; functionalism finds pain as satisfying the causal role of pain-behavior.

Functionalism has its early genealogical roots partially in behaviorism; according to the behaviorist, intelligence is the propensity to *do* or *behave* a certain way, given some set of appropriate circumstances. For soft-behaviorists like

---

[7] Ned Block, *Consciousness, Function, and Representation Collected Papers Volume I* (Cambridge, MA: MIT Press, 2007) 9, 407, 414; [hereafter: *CFR I*].

Gilbert Ryle, for instance, mental statements can be understood and made identical to a series of dispositional statements. Specifically, Ryle's position can be termed analytical behaviorism, which often is considered to be equivalent to logical behaviorism (albeit Ryle, alongside Wittgenstein, would be more aptly considered as a predecessor to logical behaviorism, proper) according to which sentences concerning mentality can be translated, without any lossage, into sentences about observable behavior. According to Ryle, Hume put forward a fallacious causal theory that one could not have a particular "idea" without having previously having had a corresponding sensation; this is, for Ryle, evidenced by Hume's so-called riddle of induction:

> [Hume] put forward the causal theory that one could not have a particular 'idea' without having previously had a corresponding sensation, somewhat as having an angular bruise involves having been previously struck by an angular object. The colors that I see in my mind's eye are, he seems to have thought, traces somehow left by the colors previously seen by me with my eyes open. The only thing that is true in this account is that what I see in my mind's eye and what I hear 'in my head' is tied in certain ways to what I have previously seen and heard. But the nature of this tie is not at all what Hume supposed.[8]

Ultimately, for Ryle, causal theories prove inadequate, with the former identified by Hume and the problem of induction. For, following the problem of induction via Hume, we experience and identify physical objects, though we cannot know laws governing their behavior. Though distinct, the issues of *whether* or *how* we identify physical objects or events, and of whether or how we identify laws governing their behavior, are correlative problems requiring conjoint solution—we can only identify physical objects and, likewise, spatio-temporal events, as objective successions by discriminating regularities in their behavior; such regularities are discerned by way of their being at least partly manifest to us by way of how they *appear* to us—that is, by distinguishing their regularities from those regularities in their appearances which depend upon our own chosen courses of action.[9] Indeed, causal relations between our surroundings and our sensory ideas simply do not suffice for explaining *how* those ideas refer to and can represent objects in our surroundings.

---

[8] Gilbert Ryle, *The Concept of Mind* (Abingdon: Routledge, 2009), 247.

[9] C.I. Lewis, *Mind and the World Order* (New York: Charles Scribner's Sons, 1929), 320.

Ryle underscores this in the aforementioned example; contra Hume, for Ryle "[t]he only thing that is true in this account is that what I see in my mind's eye and what I hear 'in my head' is tied in certain ways to what I have previously seen and heard."[10] For Ryle, mock-actions necessarily presuppose "ingenuous actions, in the sense that performing the former involves, in a special sense, the thought of the latter."[11] In turn, a person who has never seen how blue things look could not see blue things in the mind's eye (that is, they could not picture blue things in his mind's eye or recognize blue things). Thus "we learn how things look and sound chiefly and originally by seeing and hearing them."[12] Imaging, as a way of utilizing knowledge amongst many, requires the relevant knowledge has been acquired, in similar fashion. Knowledge-acquisition described as such, with it being tethered to knowledge-utilization—and thus the functional scheme writ large—demonstrates Ryle's behaviorist predilection. In this example, Ryle states that we no longer need Hume's "para-mechanical theory of traces" and all that is required for learning qua picture-thinking are those perceptual lessons which "entails some perceiving, that applying those lessons entails having learned them, and that imaging is one way of applying those lesson."[13]

Given Ryle's account, however, we have to contend with examples such as the following: the frame of mind of some person who is merely pretending to be upset, and how this is different from that of a person who genuinely is cross. For Ryle, the former's simulation involves, in some way, the "thought" of crossness. More broadly, the business of trying or simulating to behave/behavior in ways in which a cross person *would* behave is itself, at least partially, making functional use of the thought of how the person *would* behave were they genuinely cross. That is, they are using "being cross" as a prop, comporting a more or less faithful muscular representation of the cross man or woman's poutings and stampings, thus proffering the active utilization of the knowledge of how the cross man would, if genuinely cross, behave.[14]

Thus, for Ryle, there is not much between a child playing at being a pirate

[10] Ryle, *The Concept of Mind*, 247.
[11] Ibid.
[12] Ibid.
[13] Ibid.
[14] Ibid., 238-9.

and one fancying that he is a pirate. Truly, the difference is merely linguistic, tethered to our using words. However, underlying this is perhaps a more radical difference which engages with picture-thinking, a difference that occurs when and by way of the fact that we apply the words "pretend," "simulate" and "act the part," where an overt and muscular representation is given of whatever deed or condition is being "put on" while we imagine/fancy for those behaviors of mental states that are "in their heads."[15]. This brand of make-belief is what Ryle terms "imagining," "visualizing," "mind's eyes," and "going through in one's head." More important than getting rid of the superstition that picturing something is seeing a picture of it in one's "mind's eye" is the claim that epistemologists of Ryle day had, up to that point, long encouraged, asking us "us to suppose [...] that a mental picture, or a visual image, stands to a visual sensation in something like the relation of an echo to a noise."[16]

It was claimed by those espousing this stripe of epistemology that it is the case that what is occurring when I "hear," "smell," "see" corresponds something mental to that (bodily) element in perceiving. As we shall soon return, this was later embedded into Wilfrid Sellars' Myth of the Given. Consequently, Ryle considers that perceiving entails both having sensations and something else—i.e., *thinking*. Ryle details how imaging is not a function of pure sentence, or merely linguistic, and requires *learning*; what makes the imaginative operation of going through a song in one's head similar to following a heard song/rehearsing the song is in exercising humming it or playing it:

> Going through a tune in one's head is like following a heard tune and is, indeed, a sort of rehearsal of it. But what makes the imaginative operation similar to the other is not, as is often supposed, that it incorporates the hearing of ghosts of notes similar in all but loudness to the heard notes of the real tune, but the fact that both are utilizations of knowledge of how the tune goes. This knowledge is exercised in recognizing and following the tune, when actually heard; it is exercised in humming or playing it; in noticing the errors in its midperformance; it is also exercised in fancying oneself humming or playing it and in fancying oneself merely listening to it. Knowing a tune just is being able to do some such things as recognize and follow it, produce it, detect errors in the playing of it and go through it in one's head. We should not allow that a person had been unable to think how the tune went, who

[15] Ibid., 241.
[16] Ibid.

> had whistled it correctly or gone through it in his head. Doing such things is thinking how the tune goes [....] We might say that imagining oneself talking or humming is a series of abstentions from producing the noises which would be the due words or notes to produce, if one were talking or humming aloud. That is why such operations are impenetrably secret; not that the words or notes are being produced in a hermetic cell, but that the operations consist of abstentions from producing them. That, too, is why learning to fancy one is talking or humming comes later than learning to talk or hum. Silent soliloquy is a flow of pregnant non-sayings. Refraining from saying things, of course, entails knowing both what one would have said and how one would have said it.[17]

For Ryle, to consider minds and bodies as distinct from one another—as he claimed dualists did—commits to a category mistake. Cartesianism, according to Ryle, implies there is some vague abyss-like and insurmountable epistemic gap between knowledge of one's mind and the knowledge of other minds. However, despite the analytical behaviorist doctrine with which he is identified, Ryle is by no means agonistic about ontological behaviorism, as he considers behaviors, themselves, to be embedded, and thus intrinsic, with the mind/mentality. As Ryle notes, "[t]o talk of a person's mind is not to talk of a repository which is permitted to house objects that something called 'the physical world' is forbidden to house; it is to talk of the person's abilities, liabilities and inclinations to do and undergo certain sorts of things, and of the doing and undergoing of these things in the ordinary world."[18] While Ryle articulates his theory on mentality in terms of "talk about minds," it is clear that Ryle's position is that a mind is nothing but dispositions and behaviors, and nothing more than this.

While Ryle is mostly silent about specific mental occurrences , we do have some indication to his position regarding them by way of his remarks on "heed concepts," which apparently (and ambiguously) apply to mental occurrences where one may be in some apparently "heedful" state of mind with respect to behavior; to be in such a "heedful" state of mind seemingly means to be in said state without actually engaging in that behavior:

> 'heed concepts' […] refer to the concepts of noticing, taking care, attending, applying one's mind, concentrating, putting one's heart into something, thinking what one is doing, alertness, interest, intentness, studying and trying. 'Absence of

[17] Ibid., 244-5.
[18] Ibid., 199.

> mind' is a phrase sometimes used to signify a condition in which people act or react without heeding what they are doing, or without noticing what is going on.[19]
>
> Thinking or heeding what one is doing does not entail constantly or recurrently making intelligent prose moves. On the contrary, making intelligent prose moves is just one example among others of thinking or heeding what one is doing, since it is saying things, thinking what one is saying. It is one species, not the causal condition of heedful performance. But certainly didactic telling, intelligently given and intelligently received, is often an indispensable guide to execution.[20]

Despite the ambiguities and lack of precision, "heeding" may perhaps be better understood as a dispositional state, using more contemporary parlance. It also seems to be of an episodic nature:

> So while we are certainly saying something dispositional in applying such a heed concept to a person, we are certainly also saying something episodic. We are saying that he did what he did in a specific frame of mind, and while the specification of the frame of mind requires mention of ways in which he was able, ready or likely to act and react, his acting in that frame of mind was itself a clockable occurrence.[21]

In speaking of "frames of mind," Ryle's refers to those dispositional states that are "unlike motives, but like maladies and states of the weather, temporary conditions which in a certain way collect occurrences, but they are not themselves extra occurrences."[22] Ryle does not grant these "frames of mind" much significance in terms of intentionality, as he deems them temporary conditions, something akin to a short-term mental disposition that does not overdetermine (physical) causation, functionally speaking—that is, speaking in terms of mereological relations, there is no macrophysical higher level that emerges from the microphysical lower level that results in novelty, the kind of qualitative change in behavior that occurs above some "critical value" of a parameter.[23] Indeed, as

---

[19] Ibid., 118-9.
[20] Ibid., 126.
[21] Ibid., 140.
[22] Ibid., 83.
[23] "An emergent property, for Wimsatt, is a property of a system which is dependent on the mode of organization of the system's parts, how they are aggregated into the whole. Properties which are invariant against (small) changes in the modes of aggregation are non-emergent. In order for a system property P to count as emergent, P has to violate some or all of the following conditions of aggregativity:
P is invariant under rearrangements of parts of the system or replacements with 'relevantly equivalent' parts;
(ii) P is "qualitatively similar" to the property exhibited by the system when parts are added or subtracted;
(iii) P is invariant under decomposition and reaggregation of parts;

Laird Addis underscore, ontologically speaking, Ryle's "frames of mind" can be but nothing but short-term dispositions, however much, as we have seen Ryle attempts to at times avoid committing to this (or any) prescriptive theory of consciousness and the mental occurrences that make it up.[24]

Perhaps, had Ryle combined his conception of "headedness" with the aforementioned description of aping "crossness" he would have arrived at something akin to Kendall Walton's notion of "quasi-emotions," or emotions that elicit an isomorph to genuine emotions but do not stipulate the genuine belief-and-behavior pairing that genuine emotions do.[25] There is, however, a more significant problem with Ryle's conception of behavior and, more importantly, behaviorism writ large. Behaviorism, in identifying mental events/experiences as nothing beyond behavior, requires universalizing "if-then" type statements. In the case of pain, for instance, behaviorism stipulates that: "if you are in pain, then you will report it." Consider the possibility of a person who is in pain who cannot report it—they cannot wince, exclaim "ouch!" or lift their arm to relieve their pain. The behaviorist has no legitimate means of attributing such a person pain and would simply have to deny the possibility of experiencing pain for such a person. However, such people who cannot behaviorally report/communicate pain *do* exist, as in the example of selectively deafferented patients who are afflicted with pseudocoma-type "locked-in syndrome." These people cannot move, and sometimes cannot speak, but are fully conscious. Perhaps the behaviorist would rejoinder with a revised conditional account of dispositions that attempts at alleviating this burden where "if one has the ability to report their pain, then they will do so." For the behaviorist, what grounds or supervenes realized behaviors/dispositions or reports is unaccounted for.

(iv) there are no 'cooperative or inhibitory interactions" involving P among the parts of the system.' Alexander Rueger, "Physical Emergence, Diachronic and Synchronic," *Synthese* 125, no. 3 (2000): 304. William C. Wimsatt, "Reductionism, levels of organization, and the mind-body problem" in *Consciousness and the Brain*, ed. Gordon G. Globus (New York: Plenum Press, 1976), 205-67.

[24] Laird Addis, "Ryle and Intentionality," *Metaphysica* 10, no. 1 (2009): 60.

[25] E.g., "quasi-fear" can be experienced when watching a horror film, where one does not believe they are truly in danger and thus does not run out of the movie theater to avoid being murdered by the Leatherface when watching Texan Chainsaw Massacre. Kendall Walton, "Fearing Fictions," *Journal of Philosophy* 75 (1978): 5–27.

FREGE & KANT

Having carved an early ancestor vide behaviorism, and seen its weaknesses, let us now turn to functionalism proper, which has a multitudinous and variegated history; where relevant, let us also mine the relationship between functionalism and inferentialism, with an interest in the role of the particular. Specifically, we will consider Frege, Putnam and Sellars, with a bricolage carved between the first and last via the role played by inferential capacities.

Metaphysically situated, functionalism generally understood to also have an ancestor in Aristotle's teleological conception of final causation, according to which every substance in nature is considered to be conceptualizable in terms of that which realizes and "brings out" their proper essence. This would mean that a knife is not (essentially) composed of metal but, instead, by a set of functional capacities. These capacities specify the purposes for which the knife is used by way of which the knife realizes its proper role, thus revealing its essence.

Frege can be considered an early thinker of functionalism as functionalism pertains to the philosophy of language. According to Frege's conceptualization of functionalism, there is a fundamental and primary conceptual-ontological distinction between a function and an object. Specifically, Frege considered how concepts are functions of a particular type; for Frege, truth values are simply objects and "[a] concept is a function whose value is always a truth value."[26] This means that nominalizations that invoke objects also invoke functions. The generalization of functional analysis would, herewith, also extend to propositional calculus, where predicates and logical connectives were represented as linguistic functions, with names as their arguments, and denoting truth-values as their abstract referents. Frege's conception that objective representation is conceived of at the intellectual level and emphasizes the publicity and intersubjectivity of language, wherefore propositional content needs to be conceived of vide patterns of inference, thus problematizing subjectivistic methods of sense-data infallibilism and atomistic premises of representation. In both language and perception, Frege theorizes of the referent as distinct from the

[26] Gottlob Frege, "Function and Concept" in *Properties*, ed. D.H. Mellor and Alex Oliver (Oxford: Oxford University Press, 1997), 139.

mode of presentation.[27] Frege thus offers a disjunctivist lesson critical to —that there is never any specific perceptual-state kind that is commonly held between a perception of an object and a perception of another object; as Tyler Burge notes:

> Given that duplicates and illusions could make a belief false–wherever there are logically distinct events of reference—there is pressure to recognize different representational contents for different applications, even in the case of tracking that is actually successful. Where there are such possibilities, psychology will need to explain how successful tracking succeeds. To account for possible differences in veridicality among cases of contextually indiscernible duplicate switching and contextually indiscernible perceptual illusion, one must distinguish actual occurrences of context-dependent singular elements. These facts are probably the main defensible impetus for disjunctivist views. The representational contents of beliefs and perceptions involving true identities, duplicates, and illusions differ.[28]

Frege's distinction of presentation from referent here recalls Kant, as Frege and Kant both addresses issues of judgment where propositions alone are insufficient for epistemology. Before Frege, Kant had mapped conceptual explication, arguing on primarily epistemological grounds that the formal and a priori necessary conditions for the possibility of human apperception are far richer than the British empiricists had previously supposed. Kant's inventory of our cognitive capacities in the *Critique of Pure Reason* enables us to use concepts, principles and judgments cognitively. This is made particularly clear in Kant's Refutation of Idealism: Kant's critical commonsense realism about moral particulars perceived in our environs is mapped onto semantics of singular, specifically cognitive reference, which is required for experience or knowledge in any non-formal domain, such as that of spatio-temporal particulars.[29] Kant's thesis regarding the sensory manifold is that unless the contents of one's sensations have some minimum and recognizable degree of regularity and variety, perceptual synthesis does not obtain and, in turn, we cannot have the basis for what Kenneth Westphal aptly terms "cognitive judgments," using a priori or

---

[27] Gottlob Frege, "On Sense and Reference" (1892), in *Translations from the Philosophical Writings of Gottlob Frege*, eds. Peter Geach and Max Black (Oxford: Blackwell, 1966).
[28] Tyler Burge, "Disjunctivism and Perceptual Psychology," *Philosophical Topics* 33, no. 1 (2005): 34.
[29] Immanuel Kant, *Critique of Pure Reason*, trans. Paul Guyer and Allen Wood (Cambridge, UK: Cambridge University Press, 1999), B275.

empirical concepts.

> This 'affinity' (associability) of the sensory manifold is transcendental because a priori it is a necessary condition of possible apperceptive (self-conscious) human experience (of oneself as being aware of some appearances appearing to occur before, during or after others).17 It is formal because it concerns the orderliness and orderability of the contents of sensations. However, ultimately it is satisfied neither by the a priori intuitive conditions of experience (spatiality and temporality as forms of our sensory receptivity) nor by the a priori conceptual conditions of cognitive judgment. Its satisfaction is due to the 'content' or the 'object' of experience.[30]

Kant's semantics provides excellent grounds for rejecting verificationist theories of meaning, whilst insuring that genuine cognitive claims about particulars require locating them within time and space—Kant refutes the transcendent cognitive pretensions of rationalism, knowledge by acquaintance, descriptions theories of reference, and deductivist models of justification within empirical domains. Thus, Kant proves the cognitive irrelevance of merely logical possibilities to the justificatory status of empirical claims (i.e., fallibilism).[31] For Frege, sentential meaning, truth, error or approximation each requires reference to at least one relevant, extant particular. Such reference is not specified or secured by sentence meaning (intension) alone—furthermore, concepts and objects are not representations, but different types of things that our expressions refer to.

Accordingly, following Frege's conception, sensory capacities cannot in themselves represent aspects of the physical world. The representation of the objective physical environment is attainable only through the grasp of propositional structures. Consequently, it is attainable solely through a capacity for judgment. Just as perception separated from thought would not suffice for judgment, objective and shareable representation means that perception requires some supplementation from propositional judgment in order to achieve representation of an objective world. Frege's point that inferential capacities determine the logical forms of sentences and of the propositional contents of representational states. The detour through Frege (qua Kant's cognitive semantics) will prove highly relevant when we come to Sellars.

---

[30] Kenneth Westphal, *Kant's Critical Epistemology: Why Epistemology Must Consider Judgment First* (Abingdon: Routledge, 2020), §4: 35. Kant, *CPR*, A112–3, A653–4/B681–2.

[31] Westphal, *Kant's Critical Epistemology*, §15: 85.

MULTIPLE REALIZIBILITY & MACHINE-STATE FUNCTIONALISM

As it pertains to the philosophy of mind, functionalism can be traced to Aristotle's Book II of *De Anima*, in which "the soul" is defined by what allows it to survive and flourish as a perceptual, rational living being. More recently, is in the work of Putnam and Sellars where functionalism emerges as the viable alternative to the dominating phenomenalist, empiricist, and behaviorist accounts of cognition that reigned supreme. Cognitive functionalism, specifically, sets up a scenario where mental states are understood as impossible to be adequately and thoroughly characterized in terms of intrinsic identity-properties or purely in terms of overt behavior. We have already seen that the shortcomings of behaviorism, but identity-theory is what forced functionalists like Putnam and Sellars to move beyond externally demonstrated or reported behaviors.

One of those dominant accounts was the mind-brain identity theory where phenomenal concept such as pain is deemed equivalent to a material concept such as nociceptive-specific neuronal activity; as a type-b, or a posteriori, type of physicalism, mind-brain identity theory can be articulated in two modes, where the "identity" is construed as obtaining either between mental and physical types, or, when formulated more weakly, the identity obtains between mental and physical tokes. The former can be found in J.J.C. Smart's 1959 paper "Sensations and Brain Processes," which sought to polemicize behaviorism.[32] Smart argues for a novel reduction of the mind contra behaviorism, stating that mental states (e.g., pain) and phenomenal experiences (e.g., experiencing the rose's redness for the first time) can be identified with brain-states. Thus, Smart offers the "Brain Process Identity-thesis," according to which "so far as a sensation statement *is* a report of something, that something is in fact a brain process. Sensations are nothing over and above brain processes."[33]

Notably, Smart, wielding Occam's razor, lays claim to "identity" as equipollent to the intractable "is" of type-equivalency. For Smart, if mental states can truly be enumerated in terms of physical brain states, then to admit "the mental" as anything conceptually "over and above" a physical state would be unnecessary and, in turn, violate Occam's razor. This is not equipollent to eliminative materialism, as where Patricia Churchland would ask of us to

[32] J.J.C. Smart, "Sensations and Brain Processes," *The Philosophical Review* 68, no. 2 (1959), 141-156.
[33] Ibid., 145; [emphasis added].

expunge the folk-psychological term "pain" altogether, Smart merely retains it as descriptively and causally useful:

> I say that the publicly observable physical object lightning is in fact the electric discharge, not just a correlate of it. The sense-datum, or at least the having of the sense-datum, the "look" of lightning, may well in my view be a correlate of the electric discharge. For in my view it is a brain state caused by the lightning. But we should no more confuse sensations of lightning with lightning than we confuse sensations of a table with the table.[34]

In reference to afterimages, Smart remarks:

> When a person says 'I see a yellowish-orange after-image' he is saying something like this: "*There is something going on which is like what is going on when* I have my eyes open, am awake, and there is an orange illuminated in good light in front of me"[35]

Indeed, not all mental states are associated with characteristic behaviors. Such after-images, being phenomenally qualitative, are not and cannot be reliably associated with any statistically reliable characteristic behaviorist scheme; while Smart thinks the dualist may have the impulse to offer these afterimages as something fundamentally immaterial, he thinks we ought to, donning the type-B/a posteriori physicalist's cap, see these as equivalent to nothing more than brain processes/neural activity.[36] Rather than considering the afterimage as committing us to irreducibly mental conceptual states, Smart considers the way such physical states are described as articulated via "topic-neutral words," which specify mental states in terms of the stimuli that causally bring them about them. D.M. Armstrong offers a more rigorous uptake of topic-neutrality by way of logical dependence, remedying Ryle in considering the role of purpose and its causal role re: perception and belief, with topic-neutral analyses meaningfully considered for the sake of mind-brain theory.[37] Perhaps the most persuasive

---

[34] Ibid., 147.

[35] Ibid., 149.

[36] A neurosurgeon can, accordingly, observe the brain with an FMRI and "see" the experience of pain. The mind-brain identity theorist may offer that just as science has revealed lightning as electrical discharge and nothing more, water is, similarly, nothing more than H_2o. Similarly, science will reveal that yellowish-orange after-images and other phenomenal mental states and experiences are nothing more than brain states.

[37] "The logical dependence of purpose on perception and belief, and of perception." D.M. Armstrong, "The Causal Theory of the Mind." in *Neue Heft für Philosophie* 11 (1977): 88; *A materialist theory of mind* (London: Routledge, 1968).

physicalist manifestation of the mind-brain theory finds its articulation in David Lewis who, recalling his formulation, notes:

> A dozen years or so ago, D. M. Armstrong and I (independently) proposed a materialist theory of mind that joins claims of type-type psychophysical identity with a behaviorist or functionalist way of characterizing mental states such as pain [....] Positing no ambiguity without independent reason, it provides natural senses in which both madman and Martian are in pain [....] Our view is that the concept of pain, or indeed of any other experience or mental state, is the concept of a state that occupies a certain causal role, a state with certain typical causes and effects. It is the concept of a state apt for being caused by certain stimuli and apt for causing certain behavior.[38]

So follows Lewis's deductive argument for a physicalist formulation of functionalist, which begins from the scheme of functionalism to the truth of physicalism by way of identity vide some "functional specification" version of functionalism. Pain is a functionally specified state, perhaps a functionally specified brain state, according to this description, where:

1) Pain is demonstrated, via conceptual analysis, to occupy some causal role R.
2) A certain neural state will be found to be the occupant for causal role R.
3) Pain = this neural state.

Following this description, functionalism demonstrates that the *meaning* of "pain" is the same as a particular and definite description that spells out the causal role R.[39] Armstrong considered how pain and other mental states must be adequately apt for being caused in certain ways by physical stimuli plus other mental states, and thereby apt for combining with other mental states to jointly cause certain behavior. Such a mental state, like pain, is understood as the concept of a member of a system of states that together realize a pattern of causal generalizations that are pooled together to formulate some term traded in commonsense psychology, like pain. The true edge that Armstrong's conception has, however, is that the concept of pain is understood as the concept of a state that occupies a certain causal role, where whatever that state does occupies the role that is pain.

---

[38] David Lewis, "Mad Pain and Martian Pain." in *Readings in Philosophy of Psychology Vol. 1*, ed. Ned Block, (Cambridge, MA: Harvard University Press), 217-8.
[39] Ned Block, "What is Functionalism" in *CFR I*, 38.

Unlike Armstrong, Putnam and Fodor would argue from functionalism to the falsity of physicalism. Despite the two would identify that there *are* functional states/functional properties), and that mental states/properties) are identical to these functional states/properties, for Putnam and Fodor, no functional state could be reduced to a physical state. But Putnam's formulation of machine-state functionalism arguably I smore interesting that this fallacious argument against physicalism. Putnam's prolific paper, "The Nature of Mental States,"[40] was poised against behaviorism and gave the example of so-called Super-Spartans; Putnam argued, against those like Ryle, that someone may stifle one's pain-behavior, possibly to such a degree that one is not even inclined to reveal it. The argument is that behaviorism cannot account for all mental states, and it leaves out a crucial aspect of the mind, namely, our ability to experience our mental states from a first-person point of view.

Putnam's "Psychological Predicates" offers the thesis of Multiple Realizability.[41] Consequently, the question that any fully-formed functionalism after Putnam's paper must be able to answer is, thus, how to characterize the roles that define the *intentionality* of mental states and their articulation in a cognitive system. All mental states are intentional if these mental states are about or in some sense correlate to representations, including objects in the world.[42] In this regard, in its modern origins, Putnam's "machine-state" functionalism engages with the query of whether all mental states are representational: machine-state

[40] Hilary Putnam, "The Nature of Mental States" in *Philosophical Papers, Vol. II* (Cambridge, MA: 1975), 429-441.

[41] Hilary Putnam 1967, "Psychological Predicates", *Art, Mind, and Religion*, ed. W.H. Capitan and D.D. Merrill (Pittsburgh: University of Pittsburgh Press,1967), 37–48. If, as identity theory states, my experience of pain is equivalent to activity in area p in my brain, then there cannot be activity in area p without pain and there cannot be pain without activity in area p, for that is what "identity" in this context means. Thus, if the octopus brain contains nothing that resembles area p, the identity theorists would need to conclude that, octopi are unable to feel painful stimuli. Multiple Realizability offers that pain can have different neural signatures in different species, whether it be octopi or superficial functional isomorph-androids. Multiple Realization is not permitted on Smart's identity theory, as neural signature of pain shows up differently in different species, but when the neuroscientific account is complete, we will find that so-called different neural states are structurally similar enough to count as the same kind of state.

[42] G.E.M. Anscombe divides intentional states into two groups:
1) intentional states such as hopes and desires have what is referred to as a "world-to-mind direction of fit";
2) beliefs and perceptions, which aim to represent the world as it actually is; beliefs and perceptions have a "mind-to-world direction of fit."
G.E.M. Anscombe, *Intention* (Cambridge, MA: Harvard University Press: 1957), §§29–32, 51–7.

functionalism appears in the wake of computationalist approaches to the study of cognition, treating the mind as a probabilistic automaton whose operations could be understood in analogy with a Turing machine.

Putnam's functionalism models the mind on Turing machines, specifically a Probabilistic Automaton that has been generalized to allow for "sensory inputs" and "motor outputs." This Machine Table specifies, with every possible combination of a "state1 and a complete set of sensory inputs," an "instruction" that determines the probability of the next state, and accordingly the probabilities of "motor outputs."[43] Unlike behaviorism, Putnam's cognitive functionalist schema works by way of "transition properties" given by this Machine Table. As Putnam writes:

> The difficulties with 'behavior disposition' accounts are so well known that I shall do little more than recall them here. The difficulty - it appears to be more than a ' difficulty,' in fact - of specifying the required behavior disposition except as 'the disposition of X to behave as if X were in *pain*' is the chief one, of course. In contrast, we *can* specify the functional state with which we propose to identify pain, at least roughly, without using the notion of pain. Namely, the functional state we have in mind is the state of receiving sensory inputs which play a certain role in the Functional Organization of the organism.[44]

For the behavior-disposition theorist, not only are sensory inputs relevant, the Total State is relevant (thus, other values, beliefs, etc. enter into the Machine Table) and, in turn, Putnam notes that "it seems hopeless to make any general statement about how an organism in such a condition must behave; but this does not mean that we must abandon hope of characterizing the condition."[45] Such functionalist accounts provide a theory that seeks to establish formal isomorphism between the states of two systems, that preserves the relevant functional relations.[46] Putnam, in a lecture titled "The Transcendence of Reason" articulates the question of machine-state functionalism by way of the continuities

---

[43] Hilary Putnam, "The Nature of Mental States" in *Mind, Language and Reality: Philosophical Papers, Volume 2* (Cambridge, UK: Cambridge University Press, 1975), 434.
[44] Ibid., 438.
[45] Ibid.
[46] Hilary Putnam, *Reason, Truth and History* (Cambridge, UK: Cambridge University Press, 1981).

in our functional construction, asking if they play a role in our behavior while identifying behavior within a discrete set of functional possibilities:

> […] let me take up [of] whether we are computing machines or, in the more technical sense, Turing Machines. There is the question whether we are a functional organization […] some digital machine, and there is then the question, whether we can be simulated by Turing Machines. Can we do anything machines can't? I don't see any rational basis for identifying these questions [together]. There does not seem any reason to identify whether they [Turing Machines] do it [functionally behave and compute] at the psychological level in the same way [as us]. It does not seem in principle we can do anything that Turing Machines can't. This seems to be a purely physical argument [...] at the level people are talking about when asking "what can we do?" people are thinking of discrete descriptions of what we do [...] if T is a correct theory of the functioning of system 1, at the functional or psychological level, then an isomorphism between S1 and S2 must map each property and relation mentioned in T into a property and relation defined in S2 in such a way that T comes 135 out true when all references to S1 are replaced by references to S2 and all property and relation symbols are reinterpreted according to the mapping.[47]

Putnam's description recalls not only the epistemological vocabulary of Sellars but also Sellars' metalinguistic conception as, insofar as knowledge ascriptions work to normatively characterize a linguistic performance by placing it within the space of reasons, for Sellars, they function by placing a performance within a space of norms. Thus the role of a property or premise, or conclusion, is mapped onto an inferential process of justification. To say of an agent that they "know" something, Sellars claims, is to ascribe to the agent a normative status within a discursive economy. Accordingly, the epistemic facts that are tracked by the use of such normatively invested epistemic assessments are irreducible to non-epistemic facts that fulfill an empirical-descriptive role. For Sellars, when characterizing an episode or a state as that of knowing, we are not giving an empirical description of that episode or state but, instead, are placing it in the logical space of reasons, or "of justifying and being able to justify what one says."[48] Accordingly, for Sellars:

---

[47] Hilary Putnam "The Transcendence of Reason," *Howison Lectures in Philosophy* (1981). Accessed October 13, 2020. https://gradlectures.berkeley.edu/lecture/transcendence-of-reason/.

[48] Wilfrid Sellars, *Empiricism and the Philosophy of Mind* (Cambridge, MA: Harvard University Press, 1997), §36.

> […] the point is not taken care of by distinguishing between knowing bow and knowing that, and admitting that observational knowledge requires a lot to "know how." For the point is specifically that observational knowledge of any particular fact, e.g. that this is green, presupposes that one knows general facts of the form X is a reliable symptom o f Y. And to admit this requires an abandonment of the traditional empiricist idea that observational knowledge "stands on its own feet.[49]

> Thus, all that the view I am defending requires is that no tokening by S now of "This is green" is to count as "expressing observational knowledge" unless it is also correct to say of S that he now knows the appropriate fact of the form X is a reliable symptom o f Y, namely that (and again I oversimplify) utterances of "This is green" are reliable indicators of the presence of green objects in standard conditions of perception.[50]

As with Sellars' inferentialist explanation of discursive intentionality, functionalism, as articulated by Putnam, sidesteps the metaphysical question of what materially constitutes the mind, instead concerning itself with the role that mental states play as part of a system of relations. Where Putnam does defend this metaphysical question is unsatisfactory, rescinding into the crevices of internalism. Following Putnam, internalists hold that the question of what objects the world consists in "only makes sense to ask within a theory or description," with truth understood as "some sort of (idealized) rational acceptability."[51]

Contra, externalism, Putnam stakes that, pace internalism, "the world consists of some fixed totality of mind-independent objects" and, simultaneously, that "truth involves some sort of correspondence relation between words or thought-signs and external things."[52] Putnam contends that there is nothing introspectable that could "intrinsically" represent anything and, with this attempt to refute skepticism, commits to the doctrine that whatever experiences a person has, nothing of which they are aware of via introspection/reflection entails that they have *any way* to represent external objects that require a causal connection that could not exist if there were no such objects in the external world.

Putnam, in refuting the skeptical hypothesis that we are brains in a vat (BiV), states that this hypothesis "cannot possibly be true, because it is in a certain way,

---

[49] Ibid.
[50] Ibid., §37.
[51] Putnam, *Reason, Truth and History*, 49.
[52] Ibid.

self-refuting."[53] Elsewhere, Putnam proffers the stipulation (A) that:

> […] one cannot refer to certain kinds of things, e.g., trees, if one has no causal interaction at all with them, or with things in terms of which they can be described.[54]

According to Putnam, the BiV hypothesis is self-refuting because being a brain in a vat precludes being in a position to represent the vat due to the causal, and thus functional, interaction necessitated by the aforementioned claim of descriptive reference (i.e., "bracketed belief").[55] However, would not the vat create our framework conditions, such that we *would* necessarily interact with the vat? That is, the vat would contain our tractable environment, affecting us and being affected by us; thus the condition on reference that the aforementioned stipulation (A) imposes would be met. According to Putnam's conception, what must be shown is that as brains in a vat, we could not succeed in representing vats; thus, Putnam's connection between representation and causal connection implies that some necessary condition for being in the appropriate connection must be identified. The problem of reference here seems to be related to that of doxastic attitudes vis-a-vis composite reference: Putnam's case against the vat hypothesis applies without loss against the hypothesis that there are no tree. Because the concept of a tree is a composite constructed out of others concepts (i.e., the concepts of a leaf, branch, root, and so on), the causal requirement pertains to those basic concepts. In order to make stipulation (A) applicable, there must be no trees for, without trees, there could be no interaction (with trees or with things by reference to which trees can be denoted).

Thus, the metaphysical and functional levels of explanation becomes complicated. Indeed, descriptions of material systems also come to include and require functionalist explanations, with functionalism now becoming increasingly less helpful as a constraint for understanding mentality. As Ned Block argued, holding a functionalist view of the mind commits to the thesis that:

> each type of mental state is a state consisting of a disposition to act in certain ways

[53] Ibid., 7.

[54] Ibid., 16-7.

[55] Having the bracketed belief that p is being in the narrow mental state that one is in whenever one really believes that p. One can be in this mental state without really believing that p, since the causal requirements for having the concepts involved in p need not be met. Ibid., 28. Also see: Earl Conee, "Reason, Truth and History. by Hilary Putnam (review)," *Noûs* 21, no. 1 (1987): 81-95.

> and to have certain mental states, given certain sensory inputs and certain mental states. So put, functionalism can be seen as a new incarnation of behaviorism. Behaviorism identifies mental states with dispositions to act in certain ways in certain input situations desire for goal G cannot be identified with, say, the disposition to do A in input circumstances in which A leads to G, since, after all, the agent might not know A leads to G and thus might not be disposed to do A.[56]

Functionalism thus commits to two replacements in distinguishing itself from behaviorism:

i. "sensory inputs" are replaced by "sensory inputs + mental states"

ii. "disposition to act" is replaced with "disposition to act + have certain mental states."

As Block points out, functionalists wish to individuate mental states causally—since mental states have mental causes *and effects*, as well as sensory causes *and behavioral effects*, functionalists individuate mental states partly in terms of causal relations to other mental states.[57] According to Putnam's functionalism, whether a system desires something depends on whether it has internal states that have certain causal relations to other internal states (viz. to inputs and outputs). Since behaviorism makes no such "internal state" requirement, according to the functionalist, the behaviorist is guilty of liberalism—i.e., ascribing mental properties to things that do not in fact have them, as there are possible systems of which behaviorism affirms and functionalism denies that they have mental states.

Daniel Sacilotto proposes a conception of functionalism which is significantly broader, terming it "cognitive functionalism." A "cognitive functionalist" theory of mind is one which explains how a set of distinctive set of perceptual, discursive, agential, and circumspect abilities jointly instantiate the patterns of psychological and social behavior we associate with sapient cognition. This conception takes heed of the variances aforementioned and concludes that, across its different modalities, the behavior of a functional system is subject to evaluation of success or failure, such that it makes sense to speak of a possible malfunction.[58]

Similarly, Gualtiero Piccinini advocates for a Putnam-inspired "functionalism + computationalism," or computational-functionalism (also known as

[56] Ned Block, "Troubles With Functionalism," 268.

[57] Ibid.

[58] Daniel Sacilotto, *Saving the Noumenon*, ed. Nathan Brown (2020) [manuscript in preparation], §5.3.

psychofunctionalism). Piccinini begins with the concept that that the mind is the software of the brain and develops this into a thesis where the mind is but an aspect of computational organization of the brain, where computational organization means the functional organization of a computing mechanism. Accordingly, the brain is assumed to be a computing mechanism; it follows that those systems that realize minds are likened to mechanisms manipulating strings of digits according to general rules. Consequently, this interpretation cashes out between minds and many computing mechanisms, according to which minds and computing mechanisms manipulate complex combinatorial structures according to rules. Piccinini's postulated mechanism includes memory components to explain mental capacities by postulating a specific kind of mechanism with specific functional properties; memory components store programs and a processor manipulates and executes them. Computational functionalism entails that minds are multiply realizable in the sense in which different tokens of the same type of computer program can run on different kinds of hardware. According to computational functionalism, and against those foes of multiple realizability, Piccinini understands mental programs as to be specified and studied independently of how they are implemented in the brain, in the same way in which one can investigate what programs are or should be run by digital computers without worrying about how they are physically implemented. Yet Piccinini concedes that thought-experiments such as Block's China-body robot have, indeed, demonstrated that computationalism is insufficient a paradigm to explain intentionality or phenomenal consciousness. This does not entail that computationalism explains *no mental capacities* nor that intentionality cannot be explained functionally by some process other than computation.[59]

What computationalist functionalism makes clear is that an individual's mental properties can depend on external factors. More broadly, with cognitive and computationalist functionalism, a functional system is subject to evaluations of success or failure; thus we can speak of possible malfunctions in both, where an output would be something like perceptual error, misrecognition, etc. It is this potential for failure that makes functionalism attractive to explicate certain

[59] Gualtiero Piccinini, "The Mind as Neutral Software? Understanding Functionalism, Computationalism, and Computational Functionalism," *Philosophy and Phenomenological Research* 81, no.2 (2010): 273.

aspects characteristic of mental life such as cases of misperception or error, which seem difficult to explain in terms of bare behavioral regularities. Indeed, this cognitive functionalism takes into account Kant's refutation of "material" idealism, or global perceptual skepticism, vide justificatory fallibilism and Kant's cognitive semantics. Kant, who did not provide any functionalist scheme, did offer us a way to consider any Critical philosophy that deals with particular cognitive judgements. Contra Kant's cognitive semantics, infallibilism presumes that logical deduction is necessary to cognitive justification; that is, if our sole epistemological resource is logical deduction, then our method can only be conceptual analysis, which provides analytic knowledge of conceptual content—i.e., intension. However, as Kant demonstrates, logical deduction cannot provide synthetic knowledge of any truths, neither about allegedly generative causal relations between two or more things, nor about relations between our concepts and actual particulars.[60] Taking Kant's invocation of conceptual explication in his Refutation of Idealism into account by suggesting an appeal to broad conceptual necessity will not result be epistemologically efficacious unless and until we provide sufficient reliable criteria for identifying broad conceptual necessity; that is, "[u]ntil we can show how the purported broad conceptual necessities pertain to us, to our human cognitive capacities and possibilities, rather than merely to some logically possible form of nonhuman cognizance."[61] This strategy thus blocks skeptical generalization from possibility of perceptual error to the alleged universal perceptual error by identifying a pervasive dependence of human apperception upon our perceptual awareness of our natural surroundings/environs, structuring them within a causal scheme of intentionality. Contra Descartes' self-conscious awareness formula—i.e., the cogito argument—Kant's premise expounds that "I am aware of my existence as determined in time."[62] That is, each of us is aware of ourselves as being aware of some appearances happening to occur to us before, during or after other such appearances.

Kant's point is that our capacity to sort and sequence episodes of appearances-to-us requires our capacity to identify some apparent temporal

[60] Kant, *CPR*, A217/B276.
[61] Westphal, *Kant's Critical Epistemology*, 154.
[62] Kant, *CPR*, B275.

order, which relies on our success in identifying at least some persisting perceptible particulars. Sellars, like Kant, aims to block the skeptics' generalization from occasional perceptual mistake or misjudgment.[63] As Hegel insisted, it is inherent to the experience of consciousness that in representing objects in the world it distinguishes between how things appear for-itself and how things are "in-themselves," without which the possibility of falsity, nevermind truth, remains unintelligible.

Having specified this, let us distinguish that, for cognitive functionalism not to relapse into metaphysical hypostasis, it must distinguish between how functional explanation fits into the theorization of cognition in general. In this regard, Daniel Dennett famously distinguishes between three "stances" of increasing abstraction.[64] The "stances" further nuance the distinction between material and normative, or empirical and transcendental, descriptive registers of mind:

1) *the physical stance*, which studies the mind as a physical or chemical phenomenon in terms of basic mechanistic laws and a set of initial conditions;

2) *the design stance*, which studies the mind as a functional system with definite goals, orientations or tendencies, as characteristic of biological and within the theoretical purview of dynamics systems;

3) *the intentional stance*, which characterizes cognition through mentalistic explanations couched in semantic, normative and epistemic vocabularies.

It is important to note that while Dennett's framework separates intentional descriptions from the "goals" associated with functional descriptions in the "design stance," as we saw above, a normative pragmatics and inferentialist semantics can also interpret normative and semantic states functionally. By the same token, as Sellars suggested, intentional explanations can also be rendered compatible with physicalism and functionalism, as in his own forecasted "process ontological" approach that conceives of a stage in the scientific image beyond the "mechanistic" conception of causality in physics. This conception is supported by

[63] That is, from universal possibility of perceptual error or insufficient justification to the possibility of universal perceptual error or insufficient justification (global perceptual skepticism).

[64] Daniel Dennett, "Intentional Systems," *Journal of Philosophy* 68 (1971): 87-106; *The Intentional Stance* (Cambridge, MA: MIT Press, 1987).

normative concepts.[65]

Similarly, one may characterize the capacity of a being for representing their environments, by drawing a second-order analogy with theoretical reasoning. For instance, we say of the frog that it represents the fly as a light-dot within its visual field, in analogy with how a sapient agent may hold the belief that "there is a fly over there." Thus, even if intentional vocabulary is explicitly modeled on observation reports overt speech, as the Myth of Jones suggests, functional explanations to characterize natural phenomena in the physical mode are in turn implicitly modeled on the theoretical and practical reasonings which comprise the intentional stance. Let us further enumerate the Myth of Jones.

THE MYTH OF JONES

In order to illustrate the emergence of the scientific image from the manifest image, and the emergence of the manifest image from the original image in turn, in *Empiricism and the Philosophy of Mind*, Sellars famously proposes a fable of speculative anthropology, which he titles the myth of Jones. He imagines a community whose members he names our "Rylean ancestors," who describe their world and themselves while lacking the "mentalese" vocabulary of private episodes that characterizes the manifest image. These Rylean ancestors are capable not only of describing causal regularities in nature or behavioral patterns pertaining to their own doings, but also engage in "semantic talk." That is, they participate in saying that, vide overt verbal performances, they mean thus-and-so.[66]

Although the Ryleans already exhibit the capacity to undertake propositional

[65] For instance, process ontological accounts inspired by Dynamic Systems Theory, which describe sentient and sapient capacities in terms of the non-linear architecture of "self-maintaining systems."

[66] "With the resources of semantical discourse, the language of our fictional ancestors has acquired a dimension which gives considerably more plausibility to the claim that they are in a position to talk about thoughts just as we are. For characteristic of thoughts is their intentionality, reference, or aboutness, and it is clear that semantical talk about the meaning or reference of verbal expressions has the same structure as mentalistic discourse concerning what thoughts are about. It is therefore all the more tempting to suppose that the intentionality of thoughts can be traced to the application of semantical categories to overt verbal performances, and to suggest a modified Rylean account according to which talk about so-called 'thoughts' is shorthand for hypothetical and mongrel categorical-hypothetical statements about overt verbal and nonverbal behavior, and that talk about the intentionality of these 'episodes' is correspondingly reducible to semantical talk about the verbal components." Sellars, *EPM*, §50.

attitudes and use semantic vocabulary, enjoying a robust intentionality or sapient mindfulness, they nevertheless lack the intentional vocabulary required for making explicit propositional-attitude ascription. Accordingly, terms such as "meaning" and "truth" are understood as part of a "behavioristic psychology," pertaining to observable objects and events. But "genius Jones" is on the brink of a decisive theoretical breakthrough, leading to the birth of the manifest image of man in the world. Jones realizes that by postulating inner episodes as the mediating causes of overt action he can explain human behavior more robustly. These psychological, internal episodes are intro-spectible psychological episodes, which remain observable to the agents that bear them. Acts of introspection are accordingly understood as analogous to extro-spectible perceptual acts.

Such intros-pectible episodes are of two fundamental kinds: i) thoughts, and ii) sensations. Thoughts are introduced in analogy with the syntactic structure of declarative sentences and, thus, as behaviorally observed in overt speech; thoughts are, therefore, likewise semantically organized in inferential relations of incompatibility and consequence. Sensations, on the other hand, are introduced as states of perceivers; sensations are construed in analogy with the syntactic structure of perceptual reports. These perceptual reports are replicated in the mind, with their "replicas" modelled on spatially extended and colored objects observed in third-person empirical reporting's discourse. For example, having a state of sensing a red hexagon models a red hexagon inner replica which is, in turn, based on red square—the latter are conceived to be the common cause for the former. These thought-episodes and sensory impressions are critical for explaining something the original image was incapable of—the representational competences of sapient beings.

By correlating intro-spectible episodes to publicly observable objects and events, Jones can explain how there is a possible asymmetry between an agent's overt behavior (verbal and non-verbal) and states of affairs in the environment. In doing so, Jones explains instances of deception, misperception, and error. For instance, Jones can say that "Wilfrid is having a sensation of a red hexagon so he thinks there is a red hexagon over there!" upon hearing Wilfrid ejaculate: "I see a red hexagon over there!" Such is the case even when there is no red hexagon "over there." But the story does not end here. Jones' account of experience matures, and he introduces intentional locutions and propositional attitude-

ascriptions, which serve to analyze behavior dispositionally and form counterfactually robust explanatory-schemas. Thus and so, Jones can say "Wilfrid would have not hit the brakes on his car had he not believed that there was a red hexagon over there."

Such attitude-ascriptions support the ability to explain the conditions for the realization or failure of epistemic and volitional acts that bind agents to their environments. In turn, Jones' descriptive-explanatory ability provides a kind of epistemic richness when it comes to our capacity of explaining cases of error, misrepresentation, deceit, fallibility. Soon enough, Jones is able to say that "Wilfrid reached toward the hologram image of a red hexagon because he erroneously believes that there is a red hexagon over there, and he believes this because he has a sensation of a red hexagon over there even though there is no red hexagon." The emergence of the manifest image constitutes a kind of psychological and sociological theoretical breakthrough. It is through this breakthrough that experiential episodes are understood as part of a cognitive economy. This is a unique kind of inferential functionalist, which involves functionally mediating the behavior of agents. Sensations and thoughts are introduced indeed as theoretical postulates that help us to form inferences to explain the interface between the behavior of sapient agents and their environments.

A key deflationary insight here is to hold that we should separate metaphysical commitment from explanatory expediency. The intentional characterization of system that lacks mentality (e.g., Ned Block's China-body robot) should be read as serving as a provisional heuristic. That is, we can look at cases of liberalism, where mentality is attributed to those systems that do not have minds but functionally provide for the correct inputs- and out-puts noting that, when it comes to the explanatory role within a proto-theoretical framework, these schemas can, and should, eventually succeeded by a properly scientific theoretical account that dispenses of such locutions.

Sellars agrees with Kant that our commonsense spatio-temporal world of physical objects and their perceptible qualities are transcendentally ideal phenomena; for Sellars, commonsense physical objects do not exist as they are conceived within the manifest image of world—"as thus conceived, commonsense objects and events exist only in our actual or potential

representings of them."[67]

Unlike analytic empiricists and proponents of 'knowledge first epistemology', Sellars realized issues about perceptual judgment are subtle and crucial. For Sellars conceptual explication is essential but not sufficient for understanding and resolving substantive philosophical issues; conceptual analysis seeks explicit, a priori, and thus non-empirical specifications/definitions/analyses of principles by providing necessary and sufficient conditions of meaning or proper use. In contrast, conceptual explication is the partial and provisional specification of key terms in use, as explications, unlike analyses, are tied by actual linguistic practice to their relevant domains of thought, action and inquiry—therefore also to intellectual and cultural history. Much like the classical Pragmatists, Sellars explicates our concepts-in-use to gain theoretical understanding.[68]

Following Carnap, Sellars' philosophy of language focuses on semantic ascent to a constructed formalized meta-language: all abstract entities are to be defined in and confined to the meta-language—valid inferences within any language are specified in its meta-language. Thus proofs are neither more nor less than deductions which accord with the rules instituted by the meta-meta-language (e.g., Carnap's L- and P-rules).

Like Lewis, Sellars develops a distinctive conceptual pragmatism but unlike C.I. Lewis, Sellars defends 'synthetic necessary truths', necessary truths which depend upon their subject matter.[69] Like Kant and C.I. Lewis, Sellars argues "that standard empiricist vies of perception and sensory evidence are irreparably flawed."[70] A key question of Kant's First Critique is that of intentionality: how are we able to be aware of objects or events within the mind?[71] Key to intentionality is intension pertaining to individuals, universals, and states of affairs.[72] For Sellars, the actual existence of individuals and their characteristics in the world can be recognized or otherwise thought about because our states, our thoughts, and our language are structured by functional counterparts of individuals, their attributes,

---

[67] Wilfrid Sellars, *Science and Metaphysics: Variations on Kantian Themes* (London: Routledge & Kegan Paul, 1978), SM 2.29-2.32, 2.47-2/47, 2.49, 2.58-2.71, 2.71n.
[68] Sellars, *SM*, 4.24-30, .52, .58, 5.48ff, 6.7.
[69] Ibid., 2.53, 3.18-19
[70] Sellars, *EPM*, 127-196.
[71] Kant, *CPR*, A197/B242.
[72] Sellars, *SM*, 3.1-11

and our experiences of them. Sellars takes seriously Wittgenstein's notion of language games; in Sellars' metalanguage, attributes are treated as classifications of characteristics of things and individuals are treated as instances of various characteristics.

Pace Sellars, our classificatory intensions function something like Fregean senses, i.e., ways of being given, within actual or possible acts of representing; thus, characterizing these counterpart functions, Sellars treats abstract singular terms as distributive singular terms.[73]

> 'Extensions' are individuals who or which exemplify characteristics classified in intensions Sellars proposes a functional role semantics which relieves the explanatory need for exemplification; Sellars' clarificatory philosophy of mind regarding judgment contends that exemplification, like truth, 'is a matter of the semantical correctness of a certain performance—roughly the de-quoting of a quote expression.[74]

Pace Sellars' conception of picturing, semantic assertability, or "S-assertability" offers an important kind of internal- and inferential-representation (viz. i-representation) and picturing provides us with an attempt to specify a kind of environmental- and external-representation (viz. e-representation).[75] Following Sellars, Huw Price's rehabilitation of picturing/signification deals with intention qua metaphysics. The central target of that critique is Representationalism, understood as the idea that:

> There are various motivations for playing the philosophical game with restrictions of this kind, but by far the most influential, in contemporary philosophy, is the one I discussed in the previous lecture. It rests on two intuitions, or implicit assumptions. The first is a kind of prototheory about language, in the light of which the game seems to provide a useful informal model of the relation of language to the world. This proto-theory accords a key role to the idea that the function of statements is to 'represent' worldly states of affairs and that true statements succeed in doing so.[76]

To be Representationalist, as such, is to think that the proper function of

---

[73] Ibid., , 3.5203.56, 4.12-4.16.
[74] Ibid., 3.51.
[75] Ibid., 4.41n,
[76] Huw Price, *Expressivism, Pragmatism, and Representationalism* (New York: Cambridge University Press, 2013), 24.

assertoric discourse is to track how the world is e-representations are but states of a system or sub-system that systematically co-vary with states of the environment within which this system is embedded. I-representations, on the other hand, serve as connective points within an inferential nexus—thus, they service and proffer propositional content. Distinguishing between these two conceptual types of representation entails that our talk about truth and reference—that is, our semantic metavocabulary—does not (pre)establish how discourse is situationally and functionally related to the external world.

Sachs' description of Price's rehabilitation brings out that Sellars' conception of cognitive activity *in rerum natura* essentially involves mapping/map-like relations between states of some information-processing subsystem—whether an animal's brain or the CPU of some android—and detectable regularities embedded within the environment. Thus, picturing items need not be linguistic. Despite Sellars emphasizes linguistic objects in his account, this is only because language is essential to what is a uniquely human mode of picturing, and not because only languages picture. Sellars' account of picturing is a subtle elaboration of Wittgenstein's insight in the *Tractatus* that "one can only say of two objects that they stand in a certain relation by placing the corresponding referring expressions in a counterpart relation."[77] The relations amongst pictured elements cannot themselves be represented as elements within the picture.[78] The elements "within a picture must stand in counterpart relations to the relations amongst the elements of whatever is pictures" as picturing is "A relation between two relational structures, such as some worldly situation" an dour linguistic, perceptual, or conceptual representings of it. Thus, referring expressions are ineliminable and the concept of factual truth is but truth as correct picture.[79]

The Sellarsian rendering of perceptual data is not immune from critique, and perhaps Tyler Burge offers the most damning version. For representational content, understood as content with intentionality, in relativized propositions is not a veridicality condition. Differences are formulated along otherness and it is by way of sortal nouns ("natural kinds") on whose basis that we can think the relations of material incompatibility and consequence between substances. The

[77] Sellars, *SM*, 4.43.
[78] Ibid., 5.18.
[79] Ibid., 5.9

notion of "kind" comprises ability-general kinds of perceptual states which are freely shareable across individuals, freely repeatable across occasions, and serve as instances of those general kinds (i.e., actual occurrent states on particular occasions). General kinds of representational content are those that primarily figure in the laws of perceptual-state formation; occurrent, context-dependent applications are central to veridicality conditions of any perceptual state and to psychological accounts of the accuracy of inaccuracy of perceptions. Only occurrent perceptual states are accurate or inaccurate. Perception is accurate only inasmuch as an occurrent perceptual state is caused by what it is a perception of. Occurrence-based aspects can be preserved in memory and shared by different psychological modes. Representational content:

> cannot constitute the commitment of a committal state like belief. It cannot constitute the condition of success for any propositional committal state. It can be true of or false of something. It cannot be true or false. So it cannot be the full psychological representational content of a committal psychological state. It is not the commitment in the sense specified in the previous paragraph. The psychological representational content of the commitment that sets the conditions under which a committal state like belief is successful must be a condition on truth. Since committal psychological states are those that underlie explanation of all other representational psychological states, the appeal to relativized propositions cannot be a full account of the psychological representational content of any propositional psychological state. The idea that a full account of the psychological content of a belief state could appeal purely to a subpropositional representational (or psychological) content—together with the particular that the content is true of or false of in given instances of a belief—rests on a mistaken conception of committal psychological kinds.[80]

In its functional and pragmatic basis, Sellars accounts for representation as more primitive than consciousness or intentionality, even if representing is augmented by the acquisition of sensing and language. Sellars indeed provides us with a meaningful way to understand a sapient system's discursively mediated use of theoretical reasonings to describe their worlds and themselves. However, we reject Sellars' claim that reasons are the only source of epistemic warrant and that perceptual beliefs without reasons cannot be epistemically warranted—warrants

[80] Tyler Burge, "Psychological Content and Egocentric Indexes," *Blockheads!*, eds. Adam Pautz and Daniel Stoljar (Cambridge, MA: MIT Press, 2019), 52.

for perceptual belief consist partially in an individual's being in perceptual states which reliably figure in the formation of true beliefs, even if they are not reason-giving:

> An animal with visual perceptions of, and as of, bodies cannot 'make sense' of the reference. It need not have reasons. It need not know any criterion for being a body: it is enough that the animal be able to discriminate and track bodies by visual perception.[81]

Animals, children, and many adults lack reasons for their perceptual beliefs but are often warranted in having them epistemically entitled to them. To be in a contentful state just is to be in a state of a certain specific representational kind. Perceptual representational content sets veridicality conditions or, more specifically, accuracy conditions—"accuracy" discerns and distinguishes perception's conditions from truth conditions, which are propositional. The fundamental point about perceptual states is that they, and hence their representational contents, can be accurate or inaccurate, within some range of approximation. "Egocentric indexes," representing particulars that are not perceived, mark the present time, the origin or anchor of a spatial framework, or the believer—thus a believer's referring depends on the structural framework of the perception or perceptual belief[82] Consequently, epistemological warrants ought not be confined to reasons—perceptual content grounds much propositional content.

SUPERVENIENCE: FROM EPISTEMOLOGY TO METAPHYSICS

The Myth of Jones makes clear that, in the case of justified belief, any such belief cannot be what Jaegwon Kim disparagingly terms a "brute fundamental fact" unrelated to the *kind* of belief it is.[83] Thus we have two further points to make:

i. When it comes to justified belief, there must be a reason for it, a reason grounded in the factual descriptive properties of that particular belief. As demonstrated by the case of Putnam, however, if the descriptive properties are based on internalism and not reference to some particular, then we

[81] Ibid., 46.
[82] Ibid., 71-79.
[83] Jaegwon Kim, *Supervenience and Mind* (Cambridge, UK: University of Cambridge Press, 1993), 235.

are setting ourselves up for Gettier-type problems.84

ii. II) Furthermore, values, even if not reducible to facts, must be consistent with the relevant framework's fact-laden environ. Objects/particulars that are indiscernible in regard to these facts must, therefore, be considered as indiscernible in regard to value. Additionally, there must be nonvaluational reasons/grounds for the attribution of values; such reasons/grounds must be generalizable, covered by rules or norms.

Consequently, I) and II) correspond to "weak supervenience" and "strong supervenience," respectively. [85] Belief in the supervenience of value upon fact is, according to Kim, fundamental to the very concepts of value and valuation. For a concept to be significant—i.e., valuational—it must be governed by some stipulated set of impositions/criteria. These criteria cannot be arbitrary, as they must ultimately rest on factual properties, characteristics, relationships of objects and events that are held under evaluation. Thus the problem of the infinite regress, or the "infinitely descending series of valuational concepts, each depending on the one below it as its criterion of application."[86] In the spirit of Wittgenstein, Lorenz Puntel reminds of that what

> are currently presented as "theories of truth" are as a rule individual subtheories of truth. The two best known arise from the traditional distinction between the concept and the criterion of truth. Theories thematizing the criterion question are

[84] Internalism can be summarized as the thesis that: S's belief that p is justified if and only if believing that p is the attitude towards p that best fits S's evidence, where the latter is understood to depend only on S's internal mental states.

[85] Kim specifies these, writing: "A weakly supervenes on B if and only if necessarily for any x and y if x and y share all properties in B then x and y share all properties in A—that is, indiscernibility with respect to B entails indiscernibility with respect to A." "A weakly supervenes on B if and only if necessarily for any property F in A, if an object x has F, then there exists a property G in B such that x has G, and if any y has G it has R." A strongly supervenes on B just in case, necessarily, for each x and each property F in A, if x has F, then there is a property G in B such that x has G, and necessarily if any y has G, it has F. [....] The idea of strong supervenience comes to this: if St. Francis is a good man, there must be some combination of these virtues (say, honesty and benevolence) such that St. Francis has it, and anyone who has it must be a good man. This particular combination of the traits, however, need not be the only one in the base family that can 'ground' being a good man; Socrates, too, is a good man, but the virtues that he has are courage and honesty rather than honesty and benevolence. Socrates is a good man in virtue of being courageous and honest while St. Francis is a good man in virtue of being honest and benevolent. Generally speaking, a supervenient property will have alternative supervenience bases - base properties that are each sufficient for the supervening property." Kim, *Supervenience and Mind*, 58.

[86] Ibid., 236.

> conveniently termed criteriological theories of truth, and theories having the concept of truth as their subject matter, definitional theories.[87]

A "criteriological" reading, on the other hand, proposes to explain away the functional description of mindless systems in terms of bare causal regularities. The criteriological realist will paraphrase locutions like "the nose's function is to smell" into causal statements such as "scent is a consequence of the nose," where functions are understood as a select subset of the causal regularities that characterize a system's behavior.[88] This strategy appears attractive not only as a means to avoid a zealous transposition of intentionality onto mindless nature, but also insofar as it is conducive to a naturalization of intentionality that reduces reasons to causes, i.e. translating the normativity and propositional attitudes associated with sapient mindfulness to causal-material regularities amenable to naturalistic description, making no use of intentional vocabulary. However, as Jay Rosenberg argues, it is difficult to see how this reductive alternative accommodates functions within a causal story. Since many things result from the action of the nose other than the circulation of scent (for instance, nose hair prevent debris from entering), it is just not clear which consequences identify those "essential" causal links that make up the functional integrity of a system, apart from its accidental causal effects.

One would accordingly need criteria to explain the "surplus content" that allows one to discern those correlations that instantiate functional proprieties from those causal relations that contain consequents which are not considered part of a system's "proper functioning." Given this awesome difficulty, it is tempting to take the reductive strategy to an eliminativist limit, claiming that functional explanations are at best proto-scientific in character, but that they do no more than isolate those causal regularities that appear to us most difficult to characterize in purely mechanical ways (all while seeming necessary to characterize a system's behavior). An additional alternative emerges, however, in relation to the criteriological translation of function into causal explanations, one which both elides teleological characterization yet preserves the propriety of functional explanation. This entails formalizing the concept of function within a

---

[87] Lorenz Puntel, *Structure and Being: A Theoretical Framework for a Systematic Philosophy*, trans. Alan White (University Park: The Pennsylvania State University Press, 2008), §198.

[88] Jay Rosenberg, *One World and our Knowledge of it* (New York: Springer, 1980), 16-19.

broader theoretical schema, distilling its operational-mathematical core from its intentional-semantic envelopment.

By flattening cognition into a barren conception of "machinic function," one does not only elide what of sentience and sapience as it is implemented in organic-human systems must remain in order for any system to count as having a mind, blurring the line between the transcendental dimension of cognition and its biophysical modes of implementation altogether. Moreover, it also falls prey to a theoretical nihilism that, however, makes its own explanatory aims and discursive position unintelligible, yielding an iteration of what John McDowell calls "bald naturalism," which incurs on a pragmatic contradiction. For just like naturalist accounts which seek to eliminate the manifest image altogether in the name of an assimilation of man into the scientific image, "machinic-computationalist-cognitive" variants of functionalism eliminate the intentional stance. If an explanatory divide between the space of reasons and nature mystifies the former and, in turn, generates metaphysical dualism, the elimination of the framework of persons in the name of empirical function renders inexplicable how "materialism" can be understood as a method, disavowing the normatively laden project of physicalism, where notions of description, explanation, and justification are seen as implicitly joined to presenting a conceptually meaningful theory of mentality.

As Joseph Rouse insists, seconding McDowell's rejoinder to bald naturalist accounts:

> "a 'baldly naturalistic' conception of ourselves not only overreaches its empirical justification. A radically comprehensive naturalism would undermine its own intelligibility as a conception of the world. The scientific image and the understanding that it promises depend upon our capacities for conceptual understanding and its rational accountability. These very capacities for conceptual thought cannot be fully assimilated within the terms of a scientific understanding of nature."[89]

This account underscores that what "bald naturalists" presuppose ushers the achievements of the natural sciences as rationally justified while forcing any conceptual-descriptive register of understanding into a domain that outpouches

[89] Joseph Rouse, *Articulating the World: Conceptual Understanding and the Scientific Image* (Chicago: University of Chicago Press, 2015), 12.

rational accountability. We seem thus suspended between two dangers: the normativist-dualist risk of making the connection between the transcendental dimension of the space of reasons and nature unintelligible, and the reductionist-monist danger of making explanation and theorization as a whole unintelligible. Indeed, the

> rocks of Scylla loom on one side, where attempts to ground conceptual content on merely 'Given' causal or experiential impacts run aground," and, on "the other beckons the whirlpool Charybdis, in which the intralinguistic coherence of purported conceptual judgments would instead become a mere 'frictionless spinning in a void.[90]

To avoid the Scylla of humanist piety as well as the Charybdis of inhumanist triviality, Negarestani warns against a "flat functionalism" according to which "function implies an unconstrained abstract realization in so far as there is one-to-one correspondence between the realizer and realized features. Function a abstractly realizes activity b if elements of a map onto or are isomorphic with elements of b."[91] This is the danger of trivializing functional propriety by way of a "formalization" that unwittingly reiterates indiscriminate transposition. An adequate cognitive functionalism must thus theorize the invariant conditions and generic capacities which instantiate cognition in distinction from their empirical basis. Functional complexity becomes ubiquitous when any function can be realized by any kind of input-output procedure and, accordingly, any kind of "stuff," understood in the broadest sense.

The functional description of mind, requires an account of the integration of distinct processes, activities, and roles—both causal and logical. Without a precise account of this integration—how different activities with causal or logical roles are put together and integrated, and how different constraints are satisfied—the description of mind is merely the description of arbitrary "stuff." Meeting the demarcation question concerning how to isolate the functional proprieties of sapient cognition must be compatible with a naturalist perspective that answers the emergence question concerning how mind appears in the course of natural history. And this requires that we explain in what sense functional explanations can also play an explanatory role, in describing the conditions for the emergence

[90] Ibid., 224.
[91] Negarestani, *Intelligence and Spirit*, 12.

of mind in nature. We find an example of such a mature attempt in Jay Rosenberg's naturalist answer to bind teleological and criteriological readings: the "surplus content" which tracks the proper functioning of biological systems is intelligible by embedding linear causal regularities into a holistic and diachronic explanatory frame, i.e., the theory of evolution describes how processes of adaptation, heredity, selection and mutation explain the tendential behavior of organic integrity. Given our earlier example, the sought "surplus content" is an implicit appeal to the contributions of scent to the biological integrity and adaptability of organisms so structured (i.e., efficient internal transport of oxygen and information to the olfactory bulb, with odors directly routing to the limbic system, including the amygdala and hippocampus, regions related to emotion and memory). Such an appeal becomes both explicit and explanatory in the context of an evolutionary account of the origin and proliferation of organisms possessing such olfactory systems. The theory of evolution grounds functional explanations without appeal to principles different in kind from those structuring causal explanations. In the same direction, Johanna Seibt argues that the desired continuity between bare physical regularities, functional dynamics in mindless nature, and the functionality associated with the robust intentional agency of cognitive systems, can be explained within a naturalistic register in a process-ontological idiom.[92]

THE NON-REDUCTIVE PHYSICALIST'S POISONED CHALICE

Having asked about mental emergence broadly, another question that is somewhat related is still left looming: can mental states be separated from physical states? Let us return to the more innocuous initial scenario that Parfit poses in the teletransportation case, where there is no haunting overhang of my heart damage and looming death. When the aforementioned teletransportation process was activated by my pushing of a green button, my body was instantly disassembled while all the exhaustive information concerning my bodily structure and composition was recorded and transmitted to Mars. My replica and I could not be distinguished by way of appealing to any "current intrinsic physical

[92] Johanna Seibt, " How to Naturalize Sensory Consciousness and Intentionality within a Process Monism with Normativity Gradient: A Reading of Sellars" in *Sellars and His Legacy*, ed. James O'Shea (Oxford: Oxford University Press, 2016), 186-223.

differences."[93] Given that my replica is my *physical* replica, the question also is introduced of whether the replica is also my *psychological* replica? Is my replica identical with me in all mental respects? If my replica shares (aesthetic, culinary, etc.) tastes, proclivities, memories and so on, as, pace Parfit, we ought to reasonably presume that any such replica will, then this means that we have tacitly espoused the mind-body "supervenience" thesis:

> The mental supervenes on the physical in that things (objects, events, organisms, persons, and so on) that are exactly alike in all physical properties cannot differ with respect to mental properties. That is, physical indiscernibility entails psychological indiscernibility.[94]

Thus, one may say that there can be no mental difference without a physical difference. Jaegwon Kim has argued that the supervenience of mental states on physical states has dire consequences for theories of mental causation.[95]

Before moving further, let us note that when we say that there can be no mental difference without a physical difference, there is a presumption that a difference in mentality is *comprised* of physical substrates. Pain, for instance, is comprised of c-fibers firing, and we can consider what those c-fibers are made up of and go all the way down to the atoms, quarks, leptons. This mereological picture of the world and of mental states in particular describes any state as comprised of levels. According to Kim, every level consists of two components, including some set of entities "constituting the domain of particulars for that level and a set of properties defined over this domain."[96] The structural-mereological relation is as follows: entities belonging to a given layer are mereologically composed of entities belonging to the lower levels, and this relation generates a hierarchical ordering of the levels; in turn, this multitiered picture carries some idealized assumption that there is a bottom tier, a layer of entities that have no

---

[93] We say "current" to rule out the possibility of distinguishing me from my replica by way of tracing the causal chains backward to the past. We say "intrinsic" because my replica and I have different relational/extrinsic, properties—for example, we may say that I have a pet rabbit while my replica does not. Jaegwon Kim, *Philosophy of Mind* (Boulder, Westview Press: 2010), 8.

[94] Ibid., 9.

[95] Jaegwon Kim, "The nonreductivist's trouble with mental causation" in *Supervenience and Mind* (Cambridge: Cambridge University Press, 1993), 336–57.

[96] Ibid., 337.

physically significant parts.

How are the properties that characterize entities at some given level related to those entities that characterize entities of its adjacent levels? Given that entities at distinct levels are ordered by the part—whole relation, is it the case that properties associated with different levels are also ordered by some distinctive and significant relationship?

Where reductionism claims that higher-level properties are, after a certain level, reducible to lower-level ones, rivals like eliminativism and non-reductionism counter this position. Where the non-reductivist position and eliminativist position concede that higher-level properties cannot be reduced to lower-level ones, they differ on the status of those irreducible higher-level properties. Eliminativism expunges any such descriptor where a posteriori possible, following the science "all the way down." Just as Patricia Churchland thinks that using a descriptor like pain concedes to the folk-psychology speak of those misinformed psychiatrists who once believed in the descriptor "hysteria," such "useless danglers" must be expunged from the correct picture of our reality.[97]

Nonreductivism, on the other hand, holds that they can be real and genuine properties of objects and events in our world, such that there are ineliminable parts that ontologically obtain. Nonreductivism, unlike eliminativism, rejects reducibility as a means of testing whether higher-level properties are legitimate. Nonreductivism asserts that such higher-level properties can, indeed, form some autonomous domain for the sake of an independence "special science" that is, itself, ontologically irreducible to the sciences that occur and draw from lower-level phenomena. This is an attempt at respecting autonomy and the closedness of causal processes at fundamental physical levels, describing self-contained levels that are immune to causal intrusions from neighboring levels. As it relates to the mind-body problem, specifically, non-reductive physicalism articulates two characteristic theses that apply to nonreductivism:

> It consists of the two characteristic theses of nonreductivism: its ontology is physical monism, the thesis that physical entities and their mereological aggregates are all that there is; but its "ideology" is antireductionist and dualist, consisting in the claim that psychological properties are irreducibly distinct from the underlying physical

[97] Patricia Churchland, *Neurophilosophy: Toward a Unified Science of the Mind/Brain* (Cambridge, MA: The MIT Press, 1986).

> and biological properties. Its dualism is reflected in the belief that, though physically irreducible, psychological properties are genuine properties nonetheless, as real as underlying physical-biological properties. And there is a corollary about psychology: psychology is an autonomous special science independent of the physical and biological sciences in its methodology and in the concepts and explanations it generates.[98]

According to non-reductive physicalism, all concrete existents are physical and there are no nonphysical particulars or mental substances. When invoking mental states in psychological explanations, we are "presumptively committed to their causal efficacy; for any phenomenon to have an explanatory role, its presence or absence in a given situation must make a difference—a causal difference."[99] Why would Kim wish to save mental causation and treat this as something fundamentally real?

One such answer is related to human agency's being a byproduct of our mental states, including beliefs, desires, and intentions. Such states seem to have causal effects in the physical world. Additionally, the possibility of human knowledge presupposes mental causation: perception necessitates the causation of perceptual experiences and beliefs by physical objects and events around us. Reasoning, by which we acquire new knowledge and belief from the existing fund of what we already know or believe, involves the causation of new belief by old belief; more generally, causation arguably is essential to the transmission of evidential groundedness. Memory is a complex causal process involving interactions between experiences, their physical storage, and retrieval in the form of belief. If you take away perception, memory, and reasoning, you pretty much take away all of human knowledge . To move on, it seems plain that the possibility of psychology as a theoretical science capable of generating law -based explanations of human behavior depends on the reality of mental causation: mental phenomena must be capable of functioning as indispensable links in causal chains leading to physical behavior. A science that invokes mental phenomena in its explanations is presumptively committed to their causal efficacy; for any phenomenon to have an explanatory role, its presence or

[98] Kim, "The nonreductivist's trouble," 339.
[99] Jaegwon Kim, *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation* (Cambridge, MA: MIT Press, 1998), 31.

absence in a given situation must make a difference - a causal difference.

Thus, The disjunction of nomologically possible realization (consider, for instance, pain) is said to be a heterogenous disjunction for Kim. But surely these are not perfectly heterogenous, as the disjuncts of pain resemble one another insofar as they are (all) realizations of pain. disjunction of physico-chemical properties that can, compatibly with the laws of nature, realize pain. Pain is nomically coextensive with a heterogenous physico-chemical disjunction. Therefor either both pain and the disjunction are kinds, or neither pain nor the disjunction are kinds. Why? Because the kinds in question are the *projectible properties*, and one of two nomically equivalent properties is projectible if and only if the other is. That is, more specifically, Kim is talking about "objective projectability, which hinges on a notion of objective evidential support .... if there is a certain degree of evidential support for the obtaining of one property, the same degree of objective support obtains for any nomically equivalent property whether or not anyone knows about it."[100]

This picture is made more colorful and complicated by the notion of collective mentality and collective mental states, as described by Bryce Huebner:

> The term "macrocognition" is intended as shorthand for the claim that system-level cognition is implemented by an integrated network of specialized computational mechanisms, and my account of distributed cognition is equally at home in discussions of individual or collective mentality. My central hypothesis is that a plausible defense of collective mentality must take up the perspective of macrocognition to demonstrate that groups are minded in the same sense that we are.[101]

Ecosocial insect colonies, with diffuse sensory apparatuses extending in numerous directions, is one such example; consider the example of the honeybee's so-called "hive mind". The distributed information conveyed by these individual dances allows for a huge quantity of sensory information to be integrated in a way that can underwrite the goal-directed behavior of collecting nectar. What is key to the honeybee hive mind example is that a complex interaction of various-low-level representations is produced during the process of

[100] Ned Block, "Anti-Reductionism Slaps Back," *Philosophical Perspectives* 11 (1997): 114.
[101] Bryce Huebner, *Macrocognition: A theory of Distributed Minds and Collective Intentionality* (Oxford, Oxford University Press: 2013), 5.

data acquisition. But this is not just simply limited to the example of animal collective mentality. Huebner gives the example of a Crime Scene Investigation (CSI), where the collection of data begins at an emergency call center, with a call handler recording an initial representation the crime scene in real time. Consequently, the representation is passed down to a dispatch officer, who interprets it, recording and pooling the relevant information while gating off the information that is irrelevant for the dispatching officers. Investigators will then go to the scene and collect artefacts, dusting fingerprints, examining footprints, collecting biosamples (hair folicules, dead skin, saliva, etc.) and, perhaps, discarded clothing. The goal here is to distill the scene into a collection of evidential representations, including photographs, videos, clothing, fingerprints dustings; the relevant information is extracted first so as to allow for these representations to be collected for the purposes of analysis, with the eventual goal of a future prosecution:

> The processing of information by a CSI team does not depend exclusively on the architecture of the system, nor does it depend exclusively on the intentional states of the individuals that compose the team. Whether a representation will be passed along to other members of the team is determined by a complicated set of shared background assumptions, considerations about which features of the environment are likely to be salient to others, more global considerations about what sorts of information will be useful in achieving the goal of the collectivity, and facts about how data was interpreted in the past. Each of the individual investigators needs only know what they should do when they encounter particular sorts of environmental variables, yet through the interaction and coordination of these individuals, a narrative emerges that sometimes allows for the satisfaction of the collective goal of prosecution.[102]

Similarly, romantic partnerships produce distributed memories, where couples function as transactive and macrocognitive systems. Network-centric warfare also relies on a centralized control system, alongside a standard hierarchical organization, with salient information passed to centralized control systems. Macrocognitive systems also proffers in research teams and navigation crews.

But, if mental states supervene on physical states in the scenario of

[102] Ibid., 9.

psychological events, this leaves us with "too many causes of physical behavior."[103] According to mind-body supervenience, every mental event, M, has a physical cause, p* . Kim articulates the claim of mind-body supervenience as an inter-order, rather than inter-level, scheme, writing that:

> Mental properties supervene on physical properties in the sense that if something instantiates any mental property M at t , there is a physical base property P such that the thing has P at t , and necessarily anything with P at a time has M at that time.[104]

While we may think that it makes sense to say that a mental state "m" (e.g., desire to eat) causes a physical event "p" (e.g., eating), mind-body supervenience entails that every physical event (eating) also has a physical cause p* (the neural processes that comprise being hunger, or the state we refer to re: our "desire to eat")—the physical state that implements the desire. To acknowledge that mental event m, occurring at t, is a cause of physical event p but deny that p has a physical cause at t would violate the causal closure of the physical domain.[105] This results in lapsing into Cartesian interactionist dualism, which mixes physical and nonphysical events in a single causal chain. To acknowledge that p has also a physical cause, p*, at t is to invite the question:

> "Given that p has a physical cause p*, what *causal* work is left for m to contribute? The physical cause therefore threatens to exclude, and preempt, the mental cause. This is the problem of causal exclusion.[106]

The causal exclusion problem can be evaded if mental properties (or events) are identified as physical properties (or events); this means a reduction of the mental to the physical. That is to say, if we reject distinctness then there is no need to worry about exclusion, for the mental and physical do not even appear to be competing sufficient causes. Thus exclusion is supposed to be a special problem for those who accept distinctness—dualists and non-reductive physicalists, alike

---

[103] Ibid., 26.

[104] Jaegwon Kim, *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation* (Cambridge, MA: MIT Press, 2000), 23.

[105] The principle of (physical) causal closure the principle that "if you pick any physical event and trace out its causal ancestry or posterity, that will never take you outside the physical domain. That is, no causal chain will ever cross the boundary between the physical and the non-physical." Kim, *Mind in a Physical World*, 40.

[106] Ibid., 37.

According to Kim, the exclusion argument necessitates that we accept reductive physicalism; according to the doctrine of reductive physicalism, mental and other such high-level properties of the world are identified as equipollent with lower-level properties. For Kim, the exclusion argument demonstrates that any non-reductive physicalism will result in epiphenomenalism.

In response to the problems introduced by Kim's assessment, non-reductive physicalists like Karen Bennett have responded with the aim to maintain that physicalism does not need to reduce mental properties, and other high-level properties, into lower-level physical properties. Bennet's approach is termed "compatibilism."[107] This line of "compatibilist" thought can be traced back to Normal Malcolm, Stephen Yablo, Mellor, and Pereboom and Kornblith; a similar, albeit somewhat modified position is carved out in Sydney Shoemaker's work.[108]

Consider Bennett's paper, "Being Reduced."[109] Neutral concerning the a priori entailment of the mental by the physical while rejecting the metaphysical possibility of zombie worlds, Bennett believes that non-reductive physicalists should not think that mental events and properties are not identical to any physical ones. According to Bennett, we have narrow physical properties, events, and objects on the one side of our conceptual scheme, as these figure in the laws of a clearly physical science. Neurons, for instance are narrow physical objects because they figure into neuroscience. There are, simultaneously, broad physical properties, events, and objects; these are composed on the aforementioned narrow ones by means of property-forming operations such as disjunction, conjunction, and quantification. By way of the various forms of spatio-temporal, mereological and modal limit-conditions, broadly physical events are constructed from those aforementioned narrowly physical ones. Accordingly, Bennett formulates non-reductive physicalism as the view that mental states are broadly

---

[107] Note: this ought not be confused with the compatibilism of ethics

[108] Normal Malcolm, "The conceivability of mechanism," *Philosophical Review* 77 (1968): 45–72. Stephen Yablo, "Mental causation," *Philosophical Review* 101 (1992): 245–80. D.H. Mellor and Tim Crane, "There is no question of physicalism," *Mind* 99 (1990): 185–206. Derk Pereboom and Hilary Kornblith, "The metaphysics of irreducibility," *Philosophical Studies* 63 (1991): 125–45. Sydney Shoemaker, "Causality and Properties" in *Time and Cause*, ed. Peter Van Inwagen (Dordrecht: Reidel, 1980), 109-135.

[109] Karen Bennett, "Exclusion Again" in *Being Reduced: New Essays on Reduction, Explanation, and Causation*, ed. Jacob Hohwy and Jesper Kallestrup (Oxford: Oxford University Press, 2008), 281-305.

physical and not narrowly physical. This is why Bennett thinks it is a mistake for the non-reductive physicalist to claim that mental properties are not physical at all; they are physical, but not in the narrow sense. Thus and so, there is no neurological state that we can point to and accord as identical to the mental state, for this mental state is constructed out of those neurological states/properties. There is a functional characterization couched in Bennett's conception, as she picks out some narrow physical property as the realizer of a mental state. Accordingly, Bennett's non-reductive physicalism confronts the role of the mental in causation by way of over-determination. Had there been only one of the causes of the effect, the effect would have been caused by the other (and vice versa); thus, following these two counterfactuals, the effect is vacuously true. Consider the two following conditionals:

(1) if m had happened without p, e would still have happened;

(2) if p had happened without m, e would still have happened.

The real action here pertains to (2); Bennett argues that only the physicalist can say that (2) is vacuously true or false. It is vacuously true in case it is possible that p happens without m and is false if (2) is vacuously true. According to Bennett, the dualist cannot take place in this operation.

Bennett's compatibilist strategy, grossly put, is to argue that the mental and physical may be sufficiently tightly coupled that they fail to satisfy a plausible criteria for overdetermination, but without being identical. Bennett argues that the compatibilist solution is available to the non-reductive physicalist but not the dualist. Bennett outlines the defense of non-reductive physicalism by construing the exclusion problem with the following mutually exclusive claims:

i. *Distinctness*: mental and physical properties are distinct;

ii. *Completeness*: every effect has a sufficient physical cause;

iii. *Efficacy*: mental events at least sometimes cause physical events, and do so in virtue of the mental properties of the event;

iv. *Nonoverdetermination*: events are not systematically overdetermined;

v. *Exclusion*: if an effect has more than one sufficient cause, then it is overdetermined.[110]

[110] Ibid., 281.

According to this doctrine, distinctness is maintained, but exclusion is proven to not obtain. That is, there is some effect that has more than one sufficient cause but is not overdetermined. Yablo has argued that if the mental and the physical are, indeed, related as determinable and determinate then they do not overdetermine one another. One example we may consider is the relation between the color green and emerald; "being green" and "being emerald" are distinct properties. It is possible to for something to be emerald and not green; it is possible for something to be green and not emerald. It is plausible that when a driver accelerates their stopped car at a traffic light, whether that traffic light is emerald or green do not causally compete. A driver accelerates because the traffic light is green; the driver accelerates because the traffic light is emerald. In the latter case, this is because emerald is a contained as a part of being green. Anything that is emerald is necessarily green. All members of the set green include all members of the set green.

The compatibilist argument offers that, much like parts and whole do not causally compete because they are distinct while not being "wholly distinct," similarly determinates and their determinables do not causally compete because they are distinct but not "wholly distinct". Bennett's task is to show that the mental and the physical can each be sufficient causes of some effect without thereby overdetermining this effect. For Bennett, overdetermination occurs only when two counterfactuals are both non-vacuously true.[111] If one or the other counterfactual is false or only vacuously true, Bennett argues, then there is no overdetermination. According to Bennett, non-reductive physicalists should deny that the exclusion argument is a good argument for reduction while insisting that the exclusion argument is a good argument for the claim that the mental is nothing over and above the physical.

Bennett's compatibilist view is, in fact, a nuanced denial of non-overdetermination, where non-overdetermination is understood as the claim that no events are systematically overdetermined. Compatibilism therefore presents a modal analysis of overdetermination by way of using counterfactual scenarios. Bennett espouses that an effect can have multiple sufficient causes, and that this can be the case without the effect being "problematically" overdetermined. Take

[111] Ibid., 288.

causes M and P, and effect E. According to Bennett, "problematic" overdetermination occurs only if the influences of causes M and P are strongly modally independent. That is: if cause P had happened without cause M happening, effect E still would have happened; If cause M had happened without cause P, effect E still would have happened.

Christopher Devlin Brown, in response, gives the example of a firing squad shooting someone as an example of such "problematic" overdetermination.[112] Where one shooter, and the one shooter's one bullet, would do, the scenario of death by firing squad involves multiple shooters and multiple bullets. For the sake of the aforementioned scenario, let us limit the firing squad to two shooters. Let us also limit the firing squad's two shooters to shooting one bullet each into some unfortunate condemned person. One shooter is named M and the other shooter is named P. The event E is the event of shooter M and shooter P firing a bullet each into the condemned person; that event is the death of this unnamed condemned fellow. The example at hand has one minimal assumption: that the unfortunate fellow's death is causally dependent on M or P in the actual world. Now, vide David Lewis' possible worlds, let us consider a nearby counterfactual world containing only M; in this possible world, E is a consequence of M's firing. There is another nearby counterfactual world containing only P; in this counterfactual world, E occurs after P's firing. If there is a nearby counterfactual world containing only M (as just described) and there is a nearby counterfactual world containing only P (as just described), then E is overdetermined by M and P. If E had not occurred in both the counterfactual scenarios, then it would be the case that two bullets are needed to kill this condemned person and one bullet alone would e insufficient; then, in this scenario, "M's and P's firings jointly constitute a single sufficient cause and E is not problematically overdetermined."[113] If E occurs in one of the counterfactuals and not the other, then E only has one sufficient cause.

Suppose the actual-world scenario where M and P are different names for the same shooter (perhaps M is their given name and P their nickname). This is such a world where there cannot be event E without both M and P, as there is no

[112] Christopher Devlin Brown, "Exclusion Endures: How Compatibilism allows dualists to bypass the causal closure agreement" *Analysis* 79, no. 4 (2019), 487-594.
[113] Ibid., 590.

actual world containing M shooting without P shooting. Thus, it would be vacuously true that every world containing M and not P is one in which E follows M's shooting; in this most recent case, this would be vacuously true since there are no such worlds, as M and P refer to the same person.

Contending with the mind-brain identity thesis, the non-reductionist will deny that mental and "narrowly physical" properties are the identical to each other. Instead, the non-reductionist allows mental properties to be realized by different physical states. That is, the non-reductionist concedes to Multiple Realizability: pain is not identical to c-fibers firing, as Commander Data the functional isomorph could feel pain as could an alien or cyborg with different biological-chemical neural makeup (viz. silicon). This does not affect the ability of the nonreduction to use Bennett's test for non-problematic overdetermination, for as long as either "M without P is followed by E" or "P without M is followed by E" is vacuously true, then the test is passed.

The nonreductivist physicalist, per Bennett, can have events that involve high-level mental properties, or M, and low-level physical realizers, or P, where these realizers "come out as cases of non-problematic overdetermination."[114] For the non-reductive physicalist, the physical realm metaphysically determines the mental realm. This means that there are no metaphysically possible worlds that have this actual world's P properties but without the M properties. For Bennett, causal events that involve these properties will not will "pass the test for non-problematic overdetermination."[115]. There are no such worlds in which E occurs but it is not due to P realizing M.

Dualists, on the other hand, according to Bennett, are necessarily committed to the argument that there are possible worlds that copy actual-world physical properties (P) but do not copy the mental properties (M).

Brown, however, recalls a coherent dualist position that he terms "necessitarian dualism." Necessitarian dualists embrace dispositional essentialism, as they believe that properties possess nomological profiles essentially. This position he draws from Terence Horgan and Jessica Wilson, who

[114] Ibid..
[115] Ibid.

espouse the claim that P without M causes E come out as vacuously true.[116] This position espouses property dualism, the view that consciousness or its determinates are fundamental, or perfectly natural, properties. If property dualism is true, then consciousness is a fundamental ingredient in reality (or more determinate experiential properties are) and so has obvious significance in at least one sense. Furthermore, if property dualism is true, then consciousness is highly descriptively significant because it's a fundamental property, which would be confirmed through application of a bottom-up epistemology.[117]

If this is the case, then, Bennett fails at creating a solution to the exclusion problem that a non-reductionist could endorse but a dualist cannot. The necessitarian dualist accepts that all physical effects have sufficient physical causes. However, the necessitarian dualist can, in addition, also invoke Bennett's compatibilism where physical effects are non-problematically over-caused by physical properties and the nonphysical properties that accompany. Bennett, querying whether compatibilism requires physicalism, asserts that necessitarian dualism, or "metaphysical-necessitarian-dualism" is incoherent or unmotivated. Brown, however, disagrees.

Brown has two responses to this claim. First of all, if dualism is not tied to dispositional essentialism without becoming non-dualism, then it results in contradiction. Necessitarian dualism, according to Brown, is a mongrel position. Necessitarian dualism, like epiphenomenalist dualism, argues that the nonphysical is nomologically necessitated by the physical: when P is realized in any possible world containing law L, nonphysical property M is nomologically necessitated. Unlike epiphenomenalist dualism, necessitarian dualism allows for M to be an overdetermining nonphysical cause for a physical effect. This, Brown describes, is borrowed from interactionist dualism. Thus, "necessitarian dualism

---

[116] Terence Horgan, "From supervenience to superdupervenience: meeting the demands of a material world," *Mind* 102 (1993): 555–86.
material world," Mind 102 (1993): 555–86. Jessica Wilson, "Causal powers, forces, and superdupervenience," *Grazer Philosophische Studien* 63 (200): 53–77.

[117] Property Dualism qua Non-reductive Physicalism states that the psychological properties of a system are distinct from, and irreducible to, its physical properties.] For the necessitarian dualist, given the occurrence of nonphysical mental M properties by way of the occurrence of appropriate physical conditions P, it is metaphysically impossible to have P without M.

allows the mental to be causally efficacious via overdetermination."[118] Lastly, dispositional essentialism plays a role, whereby properties are individuated by way of their nomological roles (i.e., "what a property does is constitutive of its function"). The result of adding this piece to the puzzle is that every world containing P is also a world containing L, as L is necessary for P; every world that contains P and L is, thus and so, also a world containing M. Consequently, non-physical M is necessitated by physical P in every world in which P occurs.

What of Chalmers' zombie argument? As the more rigorous instantiation of Descartes' argument for mind-body interactionist dualism, this conceivability argument proffers molecule-for-molecule duplicates of us who are functional isomorphs save for the accompanying mind. The zombie argument offers that consciousness must be something that is not a physical property or set of properties in the world but is an additional component that does not arise from the organization of the forces and particles of physics. It follows that is such a scenario is possible, then physicalism is false, as physicalism requires that the physical necessitates the mental, and the mental, at its most fundamental, is physical. The necessitarian dualist cannot appeal to the zombie argument, as the necessitarian dualist says there is no possible world in which our world's physical properties would fail to generate mental properties and, thus, there is no possible zombie world. Thus, the necessitarian dualist appeals to, say, Jacksons' knowledge argument or Levine's explanatory gap argument.[119] Such arguments rely on epistemic considerations rather than modally based arguments for dualism, arguing that certain mental properties cannot be inferred from knowledge of physical properties, or that physical properties fail to explain certain mental properties.

Thus, the non-reductive physicalist is presented with a critical juncture. The exclusion problem motivates almost all of the identical premises as the causal closure argument for physicalism. If we deny a premise in the exclusion argument, the dualist will be able to counter the causal closure argument. Bennet responds to the exclusion argument by allowing "non-problematic" determination. This is precisely why the compatibilist doctrine goads the

---

[118] Brown, "Exclusion Endures," 591.

[119] Frank Jackson, "Epiphenomenal qualia," *Philosophical Quarterly* 32 (1982): 127–36; Joseph Levine, "Materialism and qualia: the explanatory gap," *Pacific Philosophical Quarterly* 64 (1983): 354–61.

physicalist into a Catch-22. Bennett's route to escape Kim's exclusion argument by prodding us into verdant grounds by way of non-reductive physicalism to invariably ends up supplying dualists with the tools to escape the causal closure argument. Accordingly, if we argue for compatibilism, we end up undercutting the argument that bolsters physicalism.

The compatibilist thinks that we can argue against the reductive physicalist's conception of Distinctness without putting these aforementioned claims together. However, according to Brown, by rejecting Nonoverdetermination, and thus overdetermining the effect that mental causes have on physical causes, physicalism is weakened to the point of fallibility. Non-reductive physicalism espouses the causal closure argument explicitly in order to embrace and bolster physicalism, while simultaneously aiming to demonstrate that the causal closure argument elides Kim's strategy of reducing the mind down to neural phenomena (or other lower-level phenomena).

The argument for the causal closure argument is, pace Brown and Papineau, as follows:

> 'every physical effect has a sufficient physical cause', mental phenomena have physical effects, and physical effects are not systematically overdetermined, so all phenomena that causally affect the physical—including mental phenomena—are themselves physical.1 So formulated, the exclusion argument and the causal closure argument are nearly indistinguishable, save that exclusion targets non-reductive physicalism and causal closure targets dualism.[120]

As Brown delineates, Kim's "exclusion argument" argues against non-reductive physicalism by way of the same logic as the causal closure argument, which works against dualism. That is, if one, or more, of the claims in the exclusion argument are denied, then the physicalist allows for the dualist to bypass what often is lauded as the strongest argument for the thesis of physicalism. Consider that a non-reductive physicalist wants to deny Nonoverdetermination so as to allow that events can be systematically overdetermined by multitudinous causes. If we were, accordingly, to deny Nonoverdermination, then the dualist could respond that it is "acceptable for physical events to be causally

[120] Brown, "Exclusion Endures," 588; David Papineau, "The causal closure of the physical and naturalism," in *The Oxford Handbook of Philosophy of Mind*, eds. Brian McLaughlin, Ansgar Beckermann and Sven Walter (Oxford: Oxford University Press, 2009), 53–65.

overdetermined by both *physical and nonphysical causes*."[121] As Brown underscores, compatibilism allows for the necessitarian dualist to unerringly claim that causal closure is not broken by physical effects, with the dualist' scheme overdetermined by both physical and nonphysical causes.

CONCLUSION

The problem Brown sketches out is that Bennett's line of compatibilism is something that not only non-reductive physicalists can take up, but also dualists. The dualist can, just like the non-reductive physicalist, attempt to attend to the "tight modal relationship between higher- and lower-level properties required by non-reductive physicalism", with the dualist can affirm that natural laws are metaphysically necessary.[122] But the affirmation of natural laws via metaphysical necessity does not illuminate what conditions, in other worlds, would have to be met for some configuration-based candidate for consciousness. Does this mean that we are better to just concede that Commander Data does not feel pain? Ought we not designate Commander Data feeling Data, or D-Pain, just as, perhaps, the cyborg feels CyborgPain, or C-Pain, while conceding that some basic structural representational make-up is shared which allows us to reasonably communicate in a given space of reasons?

If the answer to this question is yes, then we do risk eliding very real phenomenal differences for the sake of conceding co-variance in phenomenal experiences to a statistically dominant portrait of the scientific image. This might this be not only run contrary to the first-person character of phenomenal variance but to having a more accurate science. Consider that we are already aware of the possibility of color inversion, with Inverted Earth scenarios finding a cousin in empirical cases—that is, we have scientific empirical evidence for shifted spectra by way of normal color variance and between genders:

> There are a number of [sex-linked] genetic divisions in peak sensitivities in the population that are analogous to differences in blood types (in that they are genetic polymorphisms, discontinuous genetic differences coding for different types of normal perceivers) [….] These differences in peak sensitivities don't show up in common activities, but they do reveal themselves in subtle experimental situations.

---

[121] Brown, "Exclusion Endures," 588-9; [emphasis added].
[122] Brown, "Exclusion Endures," 589.

> One such experimental paradigm uses the anomaloscope (devised in the 19th Century by Lord Rayleigh), in which subjects are asked to make two halves of a screen match in color, where one half is lit by a mixture of red and green light and the other half is lit by yellow or orange light [….] Whereas one subject may see the two sides as the same in color, another subject may see them as different—e. g., one redder than the other. When red and green lights are adjusted to match orange, women tend to see the men's matches as too green or too red. Further, variation in peak sensitivities of cones is just one kind of color vision variation. In addition, the shape of the sensitivity curves vary. These differences are due to differences in macular pigmentation, which vary with 'both age and degree of skin pigmentation' […] Hence races that differ in skin pigmentation will differ in macular pigmentation. There is also considerable variation in amount of light absorption by pre-retinal structures. And this factor also varies with age. I emphasize gender, race and age to stifle the reaction that one group should be regarded as normal and the others as defective. (That would be sexism, racism or ageism.) […] My point is that the facts about variation that I have presented give us no reason at all to regard any gender, race or age as abnormal in color vision [….] Assuming that most men and women, blacks and whites, old and young have veridical color vision, two experiences can have the same representational content but different phenomenal character.[123]

Although we can restrict ourselves to biochemical neural mechanisms when considering the questions of phenomenal consciousness, our best information about *where* in the brain consciousness happens is that every conscious content is processed by the brain-area which processes that kind of information. We know that conscious contents of motion have to do with the actions in the Medial Temporal area of the visual cortex and likely involve reciprocal connections to the lower visual area. Similarly, conscious appreciation of faces has to do with activation of the fusiform face area in the bottom of the right temporal lobe. Yet we have no "place" or unified network for pinning down phenomenal consciousness. This need not mean embracing Dennett-style illusionism but understanding that P-consciousness is not an intentional property while, simultaneously, intentional differences can make a P-conscious difference.[124]

P-consciousness is experience, P-conscious properties are experiential properties, and P-conscious states are experiential states. Since P-consciousness

[123] Ned Block, "Mental Pain" in *CFR I*, 556-7, 578

[124] Ned Block, "Two Neural Correlates of Consciousness" in *Consciousness, Function and Representation* (Cambridge, MA: MIT Press, 2007), 343—363.

differences seemingly make an intentional difference, it requires that we point to those empirical experiences (tastes, smells, etc.). Thus the challenge arises for the Kantian naturalist:

> [A] realist about phenomenal consciousness should not find first-person operationalism any more palatable than third-person operationalism. Consider the reduction of phenomenal consciousness to reflective consciousness…if phenomenal consciousness is reflective consciousness, then for the pain to be phenomenally conscious is for there to be a higher-order thought about the pain, or at least some sort of cognitive state that scans it.[125]

Hence, one answer is to concede that P-conscious properties are distinct from any functional properties. This means not that their scientific picture should be, too, however, and not that the latter should supplant the former. The vagaries yet intuitive force of phenomenal-consciousness, and phenomenal experiences, given the pervasive nature of qualia-talk in both folk-psychological domains and philosophy, indicates that there is something fundamental shared across different sapient modalities.

We ought not to confuse the use of analogical modeling within conceptual-category construction and theoretical postulation with a metaphysical identification between the model and what is postulated on its basis. Sensations are modeled on the spatial and qualitative properties attributed to physical objects while enjoying counterpart attributes in relation to the latter. The same point applies to the objective determinations tracked by alethic modal vocabulary, which would be counterpart determinations to the conceptual determinations we associate with semantic inferential relations that obtain in "the space of reasons." As Robert Brandom remarks, doting on the implications of knowledge-through-inference:

> One conclusion that emerges is that the incompatibility-and- consequence relations that articulate the contents of both theoretical and observational concepts must be understood to be subjunctively robust. By engaging in inferences tracking those relations, experiencing subjects practically confront not only facts, but the lawful relations of consequence and incompatibility that make those facts both determinate and cognitively accessible. Hegel argues that we can understand the

[125] Ned Block, "Review of Daniel Dennett, *Consciousness Explained*" in *CFR I*, 129—141.

> meaning of the categorial concepts that articulate our understanding of the structure of the objective world—concepts such as object, property, fact, and law—only by understanding what we have to do to count as practically taking or treating the world as having such structure. What we must do is use singular terms and predicates to refer and describe, use declarative sentences to assert and state facts, use alethic modal vocabulary to codify laws in the form of inference licenses, and use those laws to explain facts.[126]

If we are not to simply transpose conceptual determinations into the order of scientific realism so as to claim that nature is conceptually structured, as idealists did, and do, then we must say that the forging of conceptual frameworks and theories that make use such alethic modal vocabulary constitute part of a process through which reason progressively enhances a cognitive system's representational abilities. This results in our ability grasp nature's objective *modal structure*. Underlining the diversity of language games, to the autonomy of cultures and languages, or to encroachments of anthropocentrism in the wake of posthuman intelligence, cannot alleviate this burden. As philosophizing droogs, we must make *some* critical purchase on reality, or at least attempt to. There will always be those who underscore the fetishization of reason-giving, the limits of cosmologies and world-views but, in response, the pragmatic rationalist offers that:

> that our expressions play a suitable role in reasoning is an essential, necessary element of our saying, and their meaning, anything at all. Apart from playing such a role in justification, inference, criticism, and argument, sentences and other locutions would not have the meanings appealed to and played with by all the other games we can play with language. We philosophers should be proud to acknowledge and affirm our logocentrism, but should also justify it by an account of the relations between meaning and use, conceptual content and discursive practice.[127]

What form do these so-called "rational principles" that philosophers so often refer to take? A satisfactory answer is that which Thomas Nagel gives in the appropriately titled project, *The Last Word* (2001); rational principles play a

[126] Robert Brandom, *A Spirit of Trust* (Cambridge, MA: Harvard University Press, 2019), 23.
[127] Robert Brandom, *Between Saying and Doing* (Oxford, Oxford University Press, 2008), 43.

foundational role in the inextricable bind between deliberative cognition and sapient thought, with the rational understood as that which persists in form. For instance, cinema (whether digital or analog) persists in editing, thus its rational principle is "cutting," or the montage, as Sergei Eisenstein understood early on; understanding the natural structure of the world persists in the scientific method, and such is its form.[128]

One cannot affirm or justify the obsolescence of rationality via pragmatic contradiction. Global challenges to rationality as a whole cannot but rely on what they disavow; those inferential, discursive practices of description, explanation, and theorization, which constitute the activity of reasoning. can be superseded or revised as a result of criticism and future findings. The enterprise proceeds by the operation of methods that aspire to universal validity on empirical information; to construct a rational picture of the world mean to construct one with ourselves in it, one that makes sense of these data. Empirical particulars cannot overthrow general principles but allow us to proffer conceptual content-laden qualifications. Rationality is thus not something one can withhold judgment about as one interrogates it, since the very act of interrogation implies the justificatory practices that constitute the activity of reasoning. Despite we are able to locate emergent neural phenomena, at their moment of emergence, our diachronic aspiration for concept-formation evinces an intuitive attempt to understand the functionality of sapient thought not only in psychological terms but essentially as a social and historical collective phenomenon. Recalling Hegel's apothegm, we settle into the resolution that the functionality of sapient thought can be understood richly not only in psychological terms, but essentially as a social and historical collective phenomenon, in which consciousness not only not only represents so as to progressively know an already individuated world and subjectivity, by virtue of which agents enter an interactive-dialogical space which transforms the world and themselves.

Ekin Erkan <eerkan@gradcenter.cuny.edu>

[128] Thomas Nagel, *The Last Word* (Oxford: Oxford University Press, 2001), 24.