

## THE OTHER EXPLANATORY GAP

*Abstract:* One of the driving questions in philosophy of mind is whether a person can be understood in purely physical terms. In this presentation, I wish to continue the project initiated by Donald Davidson, whose subtle position on this question has left many more perplexed than enlightened.

The main reason for this perplexity is Davidson's rather obscure pronouncements about the normativity of intentionality and its role in supporting psychophysical anomalism – the claim that there are no laws bridging our intentional states with states of our brain. Insofar as Davidson's thesis is an ontological one – about the *existence* of laws or otherwise modally significant connections between the mental and the physical – I think his critics are correct: Davidson has *not* provided us with a successful argument for psychophysical anomalism.

There is, however, a different argument, also based upon considerations about the normativity of intentionality that lead to an equally important conclusion. The conclusion is not ontological but rather epistemic: if thoughts do indeed display normativity, it is hard to understand how they would arise out of mere mechanical occurrences in the brain. To borrow a well-worn phrase, there is an “explanatory gap” between the mental and the physical. Originally coined to capture the epistemic darkness we confront in our attempt to understand phenomenal experiences in purely physical terms, the idea has yet to be explored in the area of contentful mental states or intentionality in general.

My argument shall be this: considerations about the normativity of intentionality demonstrate that there is an explanatory gap between the intentional and the physical. In fact, if there were laws of the kind Davidson denies, then the world be more mysterious than if no such laws existed. The presence of an explanatory gap explains why this is so.

### *1. Introduction: Two Explanatory Gaps*

One of the driving questions in contemporary analytic philosophy of mind is whether the mind is reducible to purely physical goings-on in one's brain. One of the more intriguing positions on this question is Davidson's anomalous monism, where reduction is possible at the level of particulars, but not at the level of properties. More specifically, anomalous monism is the idea that while each mental event is *token* identical with some physical event or other, there are no general reductive laws that link *types* of mental events with *types* of physical events. That there are no such reductive laws is known as the psycho-physical anomalism thesis. A position like anomalous monism is highly appealing because it satisfies our modern physicalist sensibilities in its commitment to the idea that all particulars are physical particulars, while at the same time,

gratifying our dualist intuition that the mental is, in some important way, distinct from the physical.

This heed to dualism is paid by the thesis of psycho-physical anomalism, a distinctly metaphysical thesis since it is about the *existence* of laws or otherwise modally significant connections between the mental and the physical. Davidson's argument for this thesis appeals to two central claims: first, intentionality is normative in that ascriptions of mental states are necessarily constrained by the norms of rationality; and second, the domain of law-governed physical events is non-normative in that there are no such norms that constrain the ascription of physical states. These two claims do not immediately lead to the conclusion there can be no laws linking the mental states with physical states, since we need a premise to the effect that the presence of a normative component in the case of intentionality and the absence of such a component in the case the physical blocks the possibility of law-like links between them. This is precisely where Davidson's argument is unfortunately vague and obscure. Along with his critics (Yalowitz 1997, Latham 1999, Tiffany 2001), I think that Davidson has *not* provided us with a successful argument for psycho-physical anomalism.

However, while Davidson may have failed to establish a metaphysical thesis, I think that his considerations about the normativity of intentionality establish a different thesis that is just as significant for the original question about the reducibility of the mental to the physical. The thesis is an epistemological one. On this thesis, statements expressing psycho-physical laws leave an "explanatory gap." A term originally coined by Joseph Levine, the idea is that reductive links between the mental and the physical may exist, for all we know, but we cannot understand *how* or *why* they exist. In this paper, I shall explain how Davidson's appeal to considerations about the normativity of intentionality in support of his argument for psycho-physical anomalism works better as an argument for the epistemic thesis that there is an explanatory gap between the intentional domain and the domain of physical facts and events. For Davidson's intuition about the normativity of intentionality and the non-normativity of physical events, furnishes a powerful articulation of *why* we are faced with an explanatory gap – why, in short, the mystery of mind-body relation is such a mystery. If I am right about this, then there are two explanatory gaps, one

pertaining to the reduction of phenomenal properties and the other pertaining to the reduction of intentional properties.<sup>1</sup>

That there may be an explanatory gap when it comes to reducing *phenomenal* properties, has long been acknowledged, ever since Levine introduced the term some 20 years ago to refashion Kripke's (1980) arguments against the identity theory as an argument about an explanatory gap.<sup>2</sup> That there may be an explanatory gap when it comes to reducing *intentional* properties or the propositional attitudes – our present concern (and the main concern of Davidson throughout his writing career) – has not been considered in the same way. Part of the reason for this, I think, has to do with different intuitions we have about our epistemic access to phenomenal properties as opposed to intentional properties. Phenomenal properties have subjective conditions of individuation; that is, whether an individual is in a phenomenal state depends solely upon the judgment of the individual. For this reason, phenomenal properties are particularly apt to invite skeptical scenarios such as inverted spectra and absent qualia, which consequently pose stumbling blocks for their materialist reduction. Such skeptical scenarios, however, don't make sense for intentional properties. That is, it is easier to conceive of phenomenal zombies than of intentional zombies. In fact, the latter just might not make any sense at all, since we tend to think that the nature of an intentional state or propositional attitude is exhausted by its causes and effects (along with the other relevant internal states), so that once the causes and effects of an intentional state are fixed, the intentional state is thereby unequivocally fixed. Intentional properties, in short, have more behaviorally ("objectively") grounded conditions of individuation than their phenomenal peers, which is why we take them to be more successfully functionalizable.<sup>3</sup> In fact, those who claim that they are working on the

---

<sup>1</sup> Under the standard taxonomy of mental states drawn by contemporary analytic philosophers of mind, mental states are divided into two basic categories: the phenomenal states and the intentional states. Phenomenal properties are the qualitative aspect of our experiences; there is something "it is like" (Nagel 1974) to have them – there is something it is like to see a shade of blue, to experience the taste of pineapple, to feel pain, and so on. Intentional properties are the semantically evaluable contents of our mental representations. While these two types of states may often go together, they are conceptually distinct and hence categorized into different taxonomic slots. Other types of mental states such as emotions, volitions, and intentions (as in having the intention to do something) may or may not be reducible to one or a composite of the former types of mental states, but clearly the phenomenal and the intentional make up much of daily mental lives.

<sup>2</sup> The notion of an explanatory gap applies to two other main anti-reductionist arguments, one by Nagel (1974) and the other by Jackson (1982).

<sup>3</sup> In addition, unlike phenomenal properties, the concepts of reference, truth, and rationality necessarily apply to their intentional peers. Intentional states have "externalist" criteria of individuation, where "external," in this

“hard problem” of consciousness typically assume that there is no troubling explanatory gap to consider in the case of intentional properties (see Chalmers 1996). The assumption that intentional properties are amenable to a materialist account is so common that they are no longer considered to be a part of the traditional mind-body problem (see Manson 2002, p.523). What is clear is that there is a confidence that intentional properties are reducible. Manson 2002 speaks for a lot of people when he says: “The contemporary mind-body problem begins with the assumption that intentionality can, in principle, be fully explained in objective causal terms. Although there is little agreement as to how we might cogently do so, this simple assumption is enough to radically transform the mind-body problem.” (p. 523)

This assumption, however, is mistaken. Intentional properties are on no better materialist footing than phenomenal ones and create just as much of an explanatory gap. The nature of this other explanatory gap shall be the main subject of this paper.<sup>4</sup> Here is how my discussion will proceed. In section 2, I present the anti-reductionist arguments given by Kripke and by Davidson; in section 3, I explain why the ontological conclusions of the arguments are better reconstrued as epistemic arguments for an explanatory gap; and section 4 concludes with a comparison of the two explanatory gaps – the one for phenomenal properties and the other for intentional properties.

---

context means facts pertaining to the individual’s surroundings that hold independently of the judgment of the individual so that the individual may fail to have privileged access to her intentional states. The asymmetry is quite dramatic. The content of one’s mental states may be *so* “external” so that the individual herself cannot know *de dicto* the content of her own thoughts (“Am I thinking of water or of twater?”). (See McKinsey 1991 for a defense of the failure of privileged access for externalist accounts of content individuation.) Such epistemic detachment makes no sense in the case of phenomenal states; one can know immediately the sensation one undergoes or the quale one enjoys.

<sup>4</sup> Horgan and Tienson 2002 call this “separatism,” which they dispute in their argument for the “interpenetration of phenomenology and intentionality.” (520) I am not here concerned about the dispute, which is about an ontological theses concerning the relationship between these two classes of mental states. My concern, rather, lies with the conceptual difference between the phenomenal and the intentional, which is all I need to motivate the possibility of two explanatory gaps.

## 2. The Conceptual Underpinnings of Kripke's and Davidson's Anti-Reductionism

There are two differences worth noting between Davidson's skepticism and that of Kripke's. First, as was already mentioned, Davidson's concern lies with *intentional* properties and propositional attitudes – mental states that have semantic and representational content – whereas Kripke's concern lies with *phenomenal* properties – the qualitative features of experiences. Second, Davidson and Kripke have different modally significant relations in mind when it comes to dealing with the reduction of their chosen targets: Kripke denies there are psycho-physical *identities*, which is consistent with the existence of psycho-physical laws, but Davidson denies that there are psycho-physical *laws*, which entails the denial of psycho-physical identities.<sup>5</sup> There are, however, two crucial similarities, and they are the concern with the reduction of the mental to the physical as *types* (properties), and a concern about the adequacy of physical *concepts* when it comes to explaining mental phenomena, a concern that is reflected in their appeal to exclusively *a priori* considerations about the differing nature of mental and physical concepts.

Here, very briefly, is Kripke's argument. Identity statements couched in terms of rigid designators are necessarily true, if true: if "*a*" and "*b*" are rigid designators and if "*a* = *b*" is true, then it is not possible for *a* and *b* to come apart (where "*a*" and "*b*" both refer to particulars or to properties). Let's now look at everyone's favorite example. Suppose that "pain" and "C-fiber stimulation" are rigid designators. The statement expressing the identity of pain and C-fiber stimulation is given as follows:

### 1. Pain is C-fiber stimulation.

The basic anti-reductionist consideration lies with the claim that can *conceive* that (1) is false. If we can conceive that (1) is false, then it is *possible* for pain and C-fiber stimulation to come

---

<sup>5</sup> While it is important to acknowledge the difference in the modal strength of the relations Davidson and Kripke aim to deny, this comparison may be misleading since Davidson and Kripke have in mind different targets – intentional properties in Davidson's case and phenomenal properties in Kripke's. In fact, it is perfectly consistent for the one to agree with the other's skeptical arguments. This is because their targets – intentional properties as opposed to phenomenal properties – are independent: one can just as well as accept or reject a nomological relation between intentional and physical properties while endorsing either an identity, a nomological relation, or no modally significant relation at all, between phenomenal and physical properties. All of which is to say that the intentional and the physical are different beasts and that they follow their own constraints.

apart. If it is possible for the two to come apart, then phenomenal properties are not reductively identifiable with any physical property, contrary to the reductive materialism in the tradition of Place 1956 and Smart 1959.

The crucial question for Kripke is this: why should we believe that the epistemic notion of conceivability can serve as a reliable guide to the metaphysically loaded notion of logical possibility? Kripke answers this by drawing the subtle but crucial distinction between correct conceivability (imaginability) and mistaken conceivability (imaginability). Here is how the distinction is drawn. All *a posteriori* reductive identifications – psycho-physical or simply physical – have an air of contingency about them. Consider the following well-established identification:

2. Heat is mean kinetic energy.

Even though we believe (2) to be true, we still sense an air of contingency, felt by the ease with which we seem to be able to conceive or imagine that (2) is false and that (3) is true.

3. Heat could have turned out to be something other than mean kinetic energy.

But, as Kripke argues, the felt contingency of true identity statements like (2) and the intuitive pull of (3) can be explained away as an illusion. If (2) is indeed true, then (3) cannot describe a logically possible world given what “heat” rigidly designates, namely, mean kinetic energy. What, then, are we thinking of when we say things like (3)? On Kripke’s account, (3) describes a world where something other than heat is causing the same sensations in us that are caused by real heat. It is a world where there is “fools-heat” – something that is qualitatively just like heat – which we mistake for real heat. This world is more appropriately captured by (4):

4. Fools-heat could have turned out to be something other than mean kinetic energy.

And (4), of course, is true. The explanation of contingency is given by deriving a statement like (4) for each statement like (3). So for each true *a posteriori* identity whose falsity we claim to conceive, there is an explanation of the apparent falsity that appeals to some fools-version of the

reduced property, a fools-version, which we mistake for the genuine article. In short, if we can identify each mistaken conceivability claim for its correctly conceivable fools-counterpart, we can use our judgments about what is conceivable as a reliable guide to logical possibility. This is Kripke's answer to the question.

We can now return to the apparent falsity of psycho-physical identities. Consider (5):

5. Pain could have turned out to be something other than C-fiber stimulation.

Can we generate a fools-counterpart to pain so that we can do for (5) what (4) did for (3)? A fools-counterpart to pain would be a sensation that feels just like pain, but isn't real pain. A moment's reflection reveals that this notion doesn't make any sense, because anything that feels just like pain simply *is* pain. Therefore, when we imagine a world where the psycho-physical identity is false, we are not talking about a logically impossible world. We are talking about a world that is indeed possible, and that is all that is needed to thwart the reduction of a mental (phenomenal) property to a physical property.<sup>6</sup>

Why is there a difference between psycho-physical reductions and inter-theoretic reductions forged in the natural sciences? Here is Kripke's diagnosis:

In the case of molecular motion and heat there is something, namely, the sensation of heat, which is an intermediary between the external phenomenon and the observer. In the mental-physical case no such intermediary is possible, since here the physical phenomenon is supposed to be identical with the internal phenomenon itself. (Kripke 1980, p. 151)

there is, as one might say, an appearance / reality distinction we must draw in the case of external phenomenal, but which does not apply to experiences (see Nagel 1974). In the case of objective, mind-independent, phenomena like heat, along with rainbows, water, lightning, and other scientific discoveries, there is a distinction one can draw between how these things appear to us, with all of the contingencies surrounding our perceptual mechanisms, and how these things

---

<sup>6</sup> Kripke's skepticism is directed upon the identification of mental properties with physical or neural properties, but it can easily be refashioned to raise a skeptical challenge to the reduction of mental properties to functional roles.

really are from a more scientifically sophisticated point of view. When we think about these things, it still makes sense to think about how these things have a nature that does not depend upon our perceptions of them. But in the case of mental phenomena like sensations and experiences, their very nature is delivered by the manner in which they are felt. Consequently, the appearance-reality distinction does not apply, as the appearance of the sensation simply is its reality.

For Davidson, the nature of the asymmetry between the intentional and the physical has to do with the presence of a *normative element* in the intentional domain and the absence of such an element in the domain of the physical. So both the phenomenal and the intentional are irreducible, but for different reasons: in the case of phenomenal properties, according to Kripke (as well as Jackson 1982 and Nagel 1974), their irreducibility comes down to the subjective-objective asymmetry; in the case of intentional properties, according to Davidson (as well as McDowell 1984 and Brandom 1994), their irreducibility comes down to the normative-non-normative asymmetry.

As with Kripke, Davidson insists upon the *a priori* nature of the argument for the irreducibility of the mental to the physical:

Nomological statements bring together predicates that we know *a priori* are made for each other – know, that is, independently of knowing whether the evidence supports a connection between the.  
- - [M]ental and physical predicates are not made for one another.  
(p. 218)

Psycho-physical generalizations are like “all emeralds are grue.” Why are we to think this? The crux of the argument lies in the different ways in which we *render intelligible* the workings of the mind as opposed to the workings of a physical system (which includes our brains and bodies). To render intelligible a range of phenomena or a system of objects, we must have a set of axioms, laws, and postulates that are constitutive of the ideas we have about the phenomena or objects whose behavior we want to understand. We need, one might say, a “linguistic framework,” a set of rules and principles that govern the ways we talk about the objects and come to understand them. Davidson calls these “constitutive principles,” as they are partly constitutive of the very concepts we use to do the intelligible rendering.

When it comes to forging law-like links between mental and physical types, this cannot be done because mental and physical concepts are governed by different constitutive principles. Mental concepts are governed by the “constitutive ideal of rationality,” where we render a person’s propositional attitudes intelligible by making them rationally coherent with each other, at least for the most part. This is what it means for intentional states to have a normative element: “[W]hen we use the concepts of belief, desire, and the rest, we must stand prepared, as the evidence accumulates, to adjust our theory in the light of considerations of overall cogency.” (1970, p. 221) Mental states are normative in that they must rationally cohere with the other mental states and behaviors attributed to an agent (all by the lights of us attributors who are aiming to understand the thoughts and behaviors of our fellow neighbors). The interstitial links between the attributed mental states must be rational in character. But no such injunction is required in our attempt to understand the workings of a physical system: “It is a feature of physical reality that physical chance can be explained by laws that connect it with other chances and conditions physically described.” (1970, p. 222) In short, we make sense of physical systems by bringing their changes under a confirmable, counterfactual-supporting, regularity, whereas we make sense of minds by bringing propositional attitudes under rules of rational coherence and consistency. We don’t make sense of physical systems by rendering rational the physical states or events that make up their behavior; such an endeavor would not make any sense since physical states do not have representational contents to which rules of rationality could be applied.

Having established the different constitutive rules for the mental and the physical, Davidson then tells us: “There are no strict psycho-physical laws because of the disparate commitments of the mental and physical schemes.” (1970, p. 222) The argument for psycho-physical anomalism is this:

1. Necessarily, if M is a mental predicate, then M is governed by the constitutive principles of rationality.
  2. Necessarily, if P is a physical predicate, then P is governed by the constitutive principles of nomological regularity.
  3. Necessarily, M and P are predicates that figure in a law only if M and P are governed by the same set of constitutive principles. (homogeneity)
  4. But M and P are governed by different sets of constitutive principles.
- 
5. Therefore, necessarily no mental predicate M and a physical predicate P can appear together in a law. (psycho-physical anomalism)

Assuming that the type of necessity is consistent throughout the argument, the argument is valid. Since (1), (2), and (3) express semantic principles, the type of necessity invoked here conceptual/logical rather than nomological/physical.

The pressing question concerns the third premise, which states that laws can feature only those predicates whose conditions of application are drawn from the same homogenous domain. What are the considerations that motivate it? Davidson seems to have two worries in mind. One is that mental predicates cannot retain their original meanings within the context of a psychophysical law: "to allow the possibility of such laws would amount to changing the subject. By changing the subject I mean here: deciding not to accept the criterion of the mental in terms of the vocabulary of the propositional attitudes." (1970, p. 216) This is the worry that I shall call the *Problem of Irrationality*. The other worry appears to be that psychophysical laws will contain unprojectible predicates given that the mental and the physical are applied according to differing rules or conventions: "In point of lawlikeness, psychophysical laws are more like 'All emeralds are grue' than like 'All emeralds are green'." (1970, p. 218) I shall call this the *Problem of Evidential Immunity*.

Let us take each of these worries in turn. What I am calling the Problem of Irrationality is inspired by Kim's 1985 attempt to make sense of this obscure third premise. The idea is this. According to some theories of inter-theoretic reduction, the reduction of one theory to another involves a form of *conceptual reduction* where we come to understand the phenomena of one

domain, say, heat, in terms of the concepts belonging to the more basic, reducing, domain, such as mean kinetic energy. Adopting this view gives us a construal according to which, to borrow Kim's 1985 apt coinage, bridge laws "transmit" the constitutive principles governing the reducing term to the term reduced:

If M is a predicate belonging to the theory that stands to be reduced and P is a predicate belonging to the reducer such that  $(x)(Px \leftrightarrow Mx)$  is a law, then the constitutive principles that govern the application of P extend their boundaries to govern the application of M.

This means that within the context of a law-like generalization like  $(x)(Px \leftrightarrow Mx)$ , the distinctly *non-rational* constitutive principles that govern the application of P would carry over to govern the application M. M would then no longer be attributed in a way that is sensitive to considerations of rationality. With no constraints of rationality imposed upon the attribution of the attitudes, we risk wide-spread irrationality, which is tantamount to abandoning the idea that we are dealing with a mind. This is why Davidson says, "[T]he nomological slack between the mental and the physical is essential as long as we conceive of man as a rational animal." (1970, p.223) The nomological reduction of mental predicates to physical predicates runs the risk of compromising the normative element of intentionality.

To avoid the Problem of Irrationality, we can maintain that law-like statements of the form  $(x)(Px \leftrightarrow Mx)$  do not transmit constitutive principles; the predicates are governed by the principles to whom they originally owe their allegiance:

If M is a predicate belonging to the theory that stands to be reduced and P is a predicate belonging to the reducer such that  $(x)(Px \leftrightarrow Mx)$  is a law M would still be governed by the rational constitutive principles of coherence and consistency and only the principles of rationality, and P would be governed by the physical constitutive principles of regularity and only the physical constitutive principles of regularity.

Here, the worry is that no knowledge of a person's brain state could ever serve as the proper kind of evidence for knowledge of his mental state (and vice versa), rendering any law-like

generalization of the form  $(x)(Px \leftrightarrow Mx)$  unconfirmable. Given that a necessary condition for projectibility is that the application of one predicate can serve as evidential warrant for the application of another, then a psychophysical law would be unconfirmable, being more like "all emeralds are grue" than "all emeralds are green."

We can spell this out more slowly this way. Suppose there are certain physical conditions that necessitate the application of P: a body of laws, together with the relevant initial conditions for a given physical state, warrant the ascription of P. And suppose there are certain other mental conditions that necessitate the application of M: a body of rational principles, together with the relevant beliefs and desires for a given mental state, warrant the ascription of M. By Davidson's lights, an ascription of P could never serve as evidence for an ascription of M because M must be warranted by relations of rational propriety, relations that are entirely absent in the case of ascribing P. Hence, neither M nor P of  $(x)(Px \leftrightarrow Mx)$  would ever be nomologically implied by the other since it would be fixed independently. In other words, if an individual believes that  $p$ , we come to know this because we come to know the other things she believes that are inferentially related to the content that  $p$ , not because we know that certain neurons are firing in such-and-such region of her cortex. Psychophysical generalizations, then, simply wouldn't count as laws – generalizations that are confirmable by their positive instances – since their antecedents would be evidentially irrelevant to their consequents.

We can now see the considerations that motivate the third premise and thus the argument for psycho-physical anomalism. Were we to countenance psycho-physical laws, we would be caught in between the horns of a dilemma: either defer to rationally insensitive physical facts to formulate mental ascriptions and risk the threat of eliminating their normative character, or hold on to the normative character of thought and divest from psycho-physical correlations their confirmability. A good look at this dilemma makes psycho-physical anomalism quite compelling.

### *3. The Retreat from the Metaphysical and the Move Towards the Epistemic*

But the inferences to these substantive metaphysical claims are by no means obvious. Consider again the Problem of Evidential Immunity. From the claim that mental ascriptions must be

evidentially warranted by whatever rational links it bears with other mental ascriptions, it is inferred that they *cannot* be evidentially warranted by physical ascriptions, leading to the conclusion that psycho-physical laws could not be confirmable. But this critical inference is nowhere supported and so patently begs the question. Surely brain states could come to serve as evidence for a mental state. We aren't anywhere near a theory that could accomplish this feat but the mere fact that mental ascriptions must be warranted by relations of rational propriety, relations that are absent in the case of ascribing a physical predicate, does not rule out this possibility. Now consider the Problem of Irrationality. From the claim that mental ascriptions would be formulated on the basis of non-rational or a-rational principles, Davidson infers that the relations between the mental ascriptions *could no longer* honor the constraints of rationality, and from this concludes that mental ascriptions cannot mean what they ought to mean under non-rational conditions of application. But again, this inference begs the question, for the claim that mental ascriptions would no longer honor the constraints of rationality within the context of a lawlike statement of the form  $(x)(Px \leftrightarrow Mx)$  is just to claim that there can be no psycho-physical laws.<sup>7</sup> In other words, mental ascriptions could be formulated on the basis of non-rational principles *and still* express rational relations among mental states. Indeed, for all we know, a mental ascription made on the basis of the constitutive principles of rationality may quite harmoniously coincide with the one that could be made on the basis of a law-like statement of the form  $(x)(Px \leftrightarrow Mx)$ .

Davidson's argument for psycho-physical anomalism, then, fails for begging the question. The same charge has long been mounted against Kripke's argument against psycho-physical identities (see Hill 1981 and 1997, Yablo 1998 and 2002). The main basis for the charge is that there is no clear way to judge whether something is coherently conceivable or not – something that Kripke himself fully acknowledges – and so when we say we can conceive that psycho-physical identities are false, we may, for all we know, be conceiving of an illusory possibility. Since we cannot trust our intuitions about what is conceivable, we cannot draw any reliable inferences about what is possible on the basis of them. Materialism, then, is not subject to falsification on conceptual grounds, contrary to the thrust of Kripke's argument.

---

<sup>7</sup> This is one of many objections raised against the argument for anomalism, and as far as I know, Brian McLaughlin (1985) has been the first to raise it. As one of many, I also think that it is the strongest, and this is the objection that I shall attempt to transform into a positive case for psychophysical laws.

However, failing to establish the metaphysical thesis that there are no modally significant psycho-physical correlations does not leave the materialist in the clear. The skeptical arguments at least succeed in shifting the burden of proof upon the materialist. Even though Davidson's argument may not succeed, it does succeed in foisting the burden of proof upon a defender of law-like statements of the form  $(x)(Px \leftrightarrow Mx)$ . For if it is correct that the mental and the physical form distinct conceptual domains, then while there is a no good reason to expect that psychophysical laws will contravene the constitutive rationality constraints, there is equally no good reason to expect that the constraints will be duly honored. Likewise in the case of evidential relevance; while there is no good reason to think that the mental and physical predicates will be evidentially irrelevant to each other, there is no positive reason to think that the necessary evidential relations will obtain. Hence, if there are psychophysical laws, then their existence must be independently defended and in a way that is harmonious with the normative character thought.

As a general matter, there are some serious epistemic difficulties with the psycho-physical correlations that a materialist should worry quite a bit about, difficulties that are powerfully brought out the arguments of Davidson as well as Kripke. Drawing this out is tantamount to a demonstration of an explanatory gap.

Let us look at how Levine redirects Kripke's argument towards an epistemic thesis about an explanatory gap. What does it mean for there to be an explanatory gap? In the context of inter-theoretic identities, the gap consists in the lack of an *a priori* entailment of the mental concept by its reducing physical concept. If materialism is true of mental properties M, then M supervenes upon physical properties P. That is, the following supervenience-conditional is true:

$$(S) (x) (Px \rightarrow Mx)$$

In a full explanation of M in terms of P, there is an *a priori* entailment of M by P. Fleshing out the details of the *a priori* entailment of the reduced by the reducer is what much of scientific progress is about. Take the case of heat. When we are convinced that we can reduce heat to mean kinetic energy, it is because we have identified 1) the causal role of heat, which is what heat does to all the things to which it is exposed (boil water, melt wax, expand metals, cause in us the sensation of warmth, etc.), and 2) the role-filler (mean kinetic energy), in such a way that

we can understand how the role-filler manages to carry out the causal role. Determining (1) and (2) is essentially the enterprise of conceptual analysis. Thus, to say that there is an explanatory gap between the reducer and the reduced is to say that there is no conceptual analysis of the reduced in terms of the reducer. According to Levine, while Kripke may not have succeeded in denying the existence of psycho-physical identities, he has nonetheless succeeded in demonstrating that there can be no conceptual analysis of the phenomenal in terms of the physical. There is no conceptual analysis because we can always ask: "why does the firing of C-fibers result in pain, as opposed to an itch or hearing middle-C." This is because no amount of examination into the nature of C-fibers will explain how or why it plays the role of pain, unlike in the case of heat and mean kinetic energy. There, the understanding of how mean kinetic energy works gives you the story about how it satisfies the causal role of heat so that once we understand mean kinetic energy, there is no further open question.

In Davidson's case, the nature of the explanatory gap is a bit different. I think the most intuitive way of arriving at it is by entertaining a little vignette. Suppose you are learning how to play a game of checkers. Not being familiar with the rules, you consult a rulebook to help plan your next move. Unbeknownst to you, the book you grab is a rulebook for chess, but it guides every move in your perfectly executed game of checkers. Now if this actually happened, it would be quite remarkable. It would be remarkable because there is no reason to expect that following the rules for chess would enable legal moves in a game of checkers (or vice versa). Indeed, there is every reason to expect that it would utterly hinder one's game of checkers.

By Davidson's lights, this is the curious situation with psycho-physical correlations: "If by absurdly remote chance we were to stumble on a ... true psychophysical generalization, we would have no reason to believe it more than roughly true" (1970, p. 216). And I gather that Davidson says this because he thinks that the governing rules of ascribing mental states differ from those concerning the ascription of physical states. Just as playing a game of checkers enjoins you to play by the rules of checkers, the enterprise of folk psychology enjoins you to ascribe mental states that rationally cohere (for the most part). And just as there are differing sets of rules for the games of checkers and chess, there are differing constitutive principles for the mental and the physical governing the ascription and interpretation of predicates belonging to the two domains. Therefore, there is no reason to expect that any given pair of mental and physical ascriptions, M and P, could be turned into a reliable, law-like, correlation since the rules

for ascribing M are not the rules for ascribing P. If M and P are governed by different constitutive principles, which means that they have different evidentiary constraints and fall to different theoretical pressures, then it really is a mystery why a given physical state should, reliably, give rise to a specific type of mental state. While the chess-checkers vignette is disanalogous in certain respects, it is effective in rendering the idea of psycho-physical laws very odd, even incredible. Just as it would be a cosmic coincidence that someone playing by the rules of checkers should manage to play legal moves throughout a game of chess, by Davidson's lights, it is a cosmic coincidence that someone trying to attribute mental states strictly on the basis of her non-rationally related neural states would end up with a system of mental states that are rationally coherent.

#### *4. Conclusion: Psycho-Physical Anomalism and the Other Explanatory Gap*

The explanatory gap, in the case of Davidson's psycho-physical anomalism, is quite similar to the one Levine attributes to the Kripkean denial of psycho-physical identities in one respect, but different in another. They are similar in that we have no idea *why* they hold; that is, for any psycho-physical supervenience conditional of the form,  $(x) (Px \rightarrow Mx)$ , we can always wonder why *that* particular conditional is true – why *that* mental type M is necessitated by *that* physical type P. They are similar, then, in that the physical facts underdetermine the mental facts. How the underdetermination is manifest, however, differs between phenomenal properties and intentional properties.

For phenomenal properties, the putative falsity of a given psycho-phenomenal supervenience conditional is intuitively motivated by the ease with which we can conjure up cases of inverted spectra and absent qualia: such and such neural stimulations may give rise to the visual sensation of green, but it could just as well have given rise to the visual sensation of red, or so it seems. For intentional properties, there is nothing quite comparable to inverted spectra or absent qualia. We can also easily conjure the falsity of psycho-intentional supervenience conditionals, but the way we can manage this is different, and here is one of the ways in which the phenomenal and the intentional differ. The belief that it is raining and the belief that it is snowing (not raining) cannot be inverted relative to each other since beliefs – all

propositional attitudes generally – are identified and individuated by their causal role in the overall network of the agent’s propositional attitudes, so that whatever occupies the role appropriate to the belief that it is raining *just is* the belief that it is raining.<sup>8</sup> This means that in the case of propositional attitudes, once a system of attitudes is in place, the beliefs one attributes relative to that system are globally constrained by all the other attitudes in the system, quite unlike the case of phenomenal states, some of which can be inverted relative to each other even after all of the other mental states – intentional and other phenomenal states – have been fixed. However, there is still an underdetermination problem for the propositional attitudes since, it is possible to derive several non-equivalent translation manuals (truth theory) for an individual that enjoin the attribution of different contents or meanings for a given individual at a time.

Mind and matter, on Davidson's picture, form two independent conceptual domains and we, as thoughtful creatures, not only have the capacity to think in terms of both, but are such that we participate in both. We are minded clumps of matter. But when we use the framework of one to describe the phenomena belonging to the other, there is the threat that we distort the nature of the subject matter we wish to understand: either we intentionalize matter or we mechanize intention. Attributing intentionality to matter is something that science has taught us to outgrow; we no longer attribute anger to thundering clouds, the act of forgiving to rainbows, or punishment with droughts. Some of our most significant scientific breakthroughs have been significant precisely because they have advanced theories that eliminate postulated intentional entities. But we cannot talk ourselves into mechanizing intention, as the deliverances of that enterprise may fail to have anything recognizably mental in it. Interestingly, this is both what the dualist as well as the eliminative materialist anticipate in the face of materialist reduction. While they disagree about the existence of irreducible mental properties, they both agree that they are irreducible. Davidson’s and Kripke’s anti-reductionist arguments furnish the vocabulary to articulate the basis for the skepticism about their reducibility.

---

<sup>8</sup> But can’t we run scenarios of content-inversion so that Oscar’s water-beliefs are switched with Twin-Oscar’s twater beliefs? Even when we recognize externalist criteria for content individuation, criteria that “slice” content broadly so that they include the goings-on of the agent’s physical environment and speech community, the qualia inversion thought experiments we can run on phenomenal properties cannot be comparably executed on intentional properties.

## References

- Chalmers, D. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.
- Chalmers, D. 2002. Consciousness and Its Place in Nature. In S. Stich and T. Warfield (eds.), *Blackwell Guide to Philosophy*. Blackwell.
- Hill, C.S. 1981. Why Cartesian Intuitions are Compatible with Identity Theories. *Philosophy and Phenomenological Research* 42:
- Hill, C. S. 1997. Imaginability, conceivability, possibility, and the mind-body problem. *Philosophical Studies* 87:61-85.
- Jackson, F. 1982. Epiphenomenal Qualia. *Philosophical Quarterly* 32: 127 – 136.
- Latham, N. 1999, Davidson and Kim on Psychophysical Laws. *Synthese* 118: 121 – 143.
- Levine, J. 1983. Materialism and qualia: The Explanatory Gap. *Pacific Philosophical Quarterly* 64: 354 – 61.
- Kim, J. 1984. Psychophysical Laws. In LePore and McLaughlin (eds.) *Actions and Events: Perspective on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell: 369-86.
- Kim, J. 1993. Can Supervenience and 'Non-Strict Laws' Save Anomalous Monism? Heil and Mele, eds. *Mental Causation*. Oxford: 19-26.
- Kripke, S.A. 1980. *Naming and Necessity*. Harvard University Press.
- Manson, N. 2002. Consciousness-Dependence and the Explanatory Gap. *Inquiry* 45: 521 – 40.
- McDowell, J. 1984. Functionalism and Anomalous Monism. In LePore and McLaughlin (eds.) *Actions and Events: Perspective on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell: 387-98.
- McKinsey, M. 1991. Anti-Individualism and Privileged Access. *Analysis* 51: 9 – 16.
- Nagel, T. 1974. What Is It Like to Be a Bat? *Philosophical Review*, 1974: 435 – 50.
- Place, U.T. 1956. Is Consciousness a Brain Process? *British Journal of Psychology* 47: 44 - 50.
- Smart, J.J.C. 1959. Sensations and Brain Processes. *Philosophical Review* 68: 141 – 56.
- Tiffany, E. C. 2001. The Rational Character of Belief and the Argument for Mental Anomalism. *Philosophical Studies* 103: 285 – 314.
- Yablo, S. 1998. *Textbook Kripkeanism and the open texture of language*. *Philosophical Quarterly*.
- Yablo, S. 2002. Coulda, woulda, shoulda. In T. Gendler & J. Hawthorne, eds. *Conceivability and Possibility*. Oxford University Press.
- Yalowitz, S. 1997. Rationality and the Argument for Anomalous Monism. *Philosophical Studies* 87: 235 – 258.