Intelligibility and the CAPE: Combatting Anti-Psychologism about Explanation (Draft)
Jonathan Waskan

1 OVERVIEW

Much of the philosophical discussion of explanations has centered around two broad conceptions of what sorts of 'things' explanations are, descriptive and objective. Proponents of each agree upon one thing: Psychology can contribute little to the study of explanations. They attempt to show this by pointing to cases of explanation where the commonly associated phenomenology of explanation (CAPE) (e.g., feelings of insight or understanding) is absent and cases where the CAPE is present without any explanations. All such arguments improperly exploit the ambiguity of 'explanation', but they do contain a kernel of truth. The CAPE is, in fact, not constitutive of explanation, not even in the oft-overlooked (third) psychological sense of the term. What appears to be essential is that one finds a happening intelligible. Here I propose a model of the psychological underpinnings of intelligibility and, ultimately, of what explanations are (in the psychological sense). I close by outlining how the psychological study of intelligibility may actually help to reveal the origins of all three concepts of explanation and, in turn, the origins of the judgments about explanation that have in large measure driven philosophical theorizing on the subject.

2 ECHOES OF HEMPEL

The study of explanations became a going concern in Anglo-American philosophy in the mid-20th century logical positivism was on the decline. To paint with broad strokes, positivistic philosophers of science sought to brand science as an institutional source of knowledge that is (unlike religion, astrology, and large swaths of philosophy) devoid of meaningless metaphysical speculations. The major sticking point was claims about theoretical goings-on (e.g., electrons), but positivists maintained that these claims could be straightforwardly 'reduced,' using the tools of formal logic, to claims about the observations that would confirm them. But there is, of course, much more to theoretical claims than meets the eye. They are typically linked in complex ways to, among other things, lots of other theoretical claims. The upshot is that even if a theoretical claim about a system were accurate, altering the system in a certain way would still not always yield the expected outcome. Theoretical claims make, it came to be acknowledged, do make assertions about happenings unseen, and in this way they also serve an important retrospective function: They provide explanations for how and why observed happenings unfold the way they do.

2.1 Hempel's Anti-psychologism

Carl Hempel's seminal work on explanation appeared in the wake of the positivistic obsession with the prediction of observables. According to Hempel (1965), science is the product of two basic human motives. One is mankind's practical desire to improve his life through foresight and control over nature. Here predictive leverage is clearly quite important. The other is to be found in "his sheer intellectual curiosity, in his deep and persistent desire to know and to understand himself and his world" (333). This requires explanations. Hempel's own interest in explanation stems from a similar motive, from his deep and persistent desire to know and to understand, "What is the nature of the explanations empirical sciences can provide? What understanding of empirical phenomena do they convey?" (333). This is what might be called an *ontological motive*. According to Hempel, it could be served by examining the "form and function" of the various kinds of explanation provided by science (333). In addition, like most

philosophers of science, Hempel views the scientific method of proposing and evaluating explanations as an exemplary way of gaining knowledge of the world. Determining how, on its best day, science determines whether an explanation is "true" or, as he puts it, merely "potential" (338) – for Hempel all genuine explanations are accurate – serves what might be called his *epistemological motive*.

Hempel undertook to study explanation by abstracting by abstracting from people engage in the act of explaining happenings to one another and from the psychological states that may result therefrom. In the footsteps of his positivist predecessors, he would abstract as well from the specific contents involved and try to discern the bare logical form of explanations utilizing the tools of formal logic. Hempel believed that just as major advances were made in proof theory through a similar double abstraction, major advances in the study of explanation, and science more generally, can be expected in this way as well. Hempel (1965) purports that a major success story for this approach was the demonstration that theoretical claims can be Ramsified (414-5). In this way, it was hoped, one could acknowledge the full complexity of theoretical claims and their essential reference to matters unseen and supply a reconstructions of their meanings in observational and paralogical terms. This achievement was made possible, says Hempel, "only by reference to a precisely formulated, and to some extent schematic, conception of scientific explanation" (415).

In any case, the eventual outcome of Hempel's double abstraction was covering-law model of explanation. On this view, an explanation is an argument, which comprises statements describing one or more laws and (in some cases) particular conditions which imply, in ways that can be captured by the principles of inductive and deductive logic, a statement describing the target happening. In this way, an explanation shows that the happening "*was to be expected,"* and thereby it "enables us to *understand why* the phenomenon occurred" (1965, 337).

Given that Hempel takes understanding to be the driving motive behind the search for explanations, one might think that in this respect he is committed to some form of psychologism about explanations.[1] Hempel was, however, careful to distinguish between genuine understanding and the mere sense, or feeling, of understanding. He claims, "it is important to distinguish here understanding in the psychological sense of a feeling of empathic familiarity from understanding in the theoretical… sense of exhibiting the phenomenon to be explained as a special case of some general regularity" (1965, 256-7). This distinction does, in fact, seem to reflect our common ways of using the term 'understanding.'[2] In one sense, 'understand' is a success verb much like (a sense of) 'see' – that is to say, in order to do it you must be successful at it. We might say, for instance, that whereas alchemists *felt* that they understood combustion, chemists do *genuinely* understand combustion. Hempel identifies the mere feeling of understanding, which as we saw has no success requirement, with feelings of empathy or familiarity. These psychological events, he claims, are neither necessary nor sufficient for genuine understanding or genuine ('theoretical' or 'scientific') explanation (257, 430). Whether or not something is an explanation does not track these subjective psychological states. Something can be an explanation in the absence of these feelings, for sometimes an explanation hinges upon an appeal to something unfamiliar, such as when the law of gravitation is used to

[1] In fact, I am inclined to think it does (de Regt 2009; also see Waskan 2006), but that is not where I wish to pitch my flag in this essay.

[2] There are, of course, other senses of 'understanding' which may or may not have little to do with explanation, as in "Eliza does not understand a word of English."

explain the behavior of falling bodies (257, 431). And something can evoke these feelings without constituting an explanation (257). For instance, that the Scholastic account of the behavior of falling bodies appealed to an object's inner desire to seek it's natural resting place. This clearly appeals to something familiar, but it is, according to Hempel, no explanation at all.

This line of argument is meant to serve Hempel's ontological motive, his desire to discern what he calls the "essential characteristics" of explanations (1965, 245). Given that Hempel equates the feeling of understanding with feelings of empathy or familiarity, we may put the central premises of his *ontological anti-psychologism* argument as follows:

> Psychologistic theories equate explanations with the feeling of understanding (viz., feelings of familiarity or empathy). Yet there are clear-cut instances of (descriptive) explanation without such feelings, and there are clear-cut instances of such feelings without explanation

The clear-cut instances in question are all descriptive in nature.

To reinforce his anti-psychologism, Hempel adds the following: "Besides, the extent to which an idea will be considered as familiar varies from person to person and from time to time, and a psychological factor of this kind certainly cannot serve as a standard of assessing the worth of a proposed explanation" (1965, 258). This concerns the epistemological motive of determining how we ought to evaluate explanations. We might re-express the central premise of Hempel's *epistemic anti-psychologism* argument roughly as follows:

> Whether or not a description counts as a good explanation is not determined by what subjects feel.

Hempel's proposal that explanations are sets of descriptions that bear logical relations to one another would be rejected by many for, among other things, its apparent inability to properly sort cases of genuine explanation from non-explanations. Better withstanding the test of time, however, are his anti-psychologistic arguments and the related assumptions that genuine understanding is distinct from the mere feeling of understanding and that genuine explanation requires accuracy. One hears echoes of them, for instance, in the works of those who favor more mechanism-based theories of explanation.

2.2 Salmon's Anti-psychologism

On mechanistic theories, explanation has not to do not with laws (at least not primarily) but with how the parts of a system collectively conspire to produce the happening of interest. The mechanistic approach was pioneered in large part by Wesley Salmon, who seems to have co-opted the aforementioned arguments and assumptions in order to defend what he called an "ontic" conception of what, in the most general sense, explanations are. The ontic conception was meant as an alternative to Hempel's "epistemic" conception which took explanation to render happenings expectable. In order to understand Salmon's anti-psychologistic, pro-ontic arguments, we must first get clear on what the ontic conception amounts to. Complicating matters, Salmon uses 'ontic' in two very different ways.

2.2.1 What is the ontic conception?

Something that surely motivated Salmon's introduction of the ontic conception is the fact that we often talk of *the* explanation for a happening, as in, "the explanation for combustion is oxidation." As Craver rightly notes, in this sense of the term, explanations are not the sorts of

things that are good or bad, right or wrong; "they just are" (2007, 27). Craver appears to view this as the right way to understand Salmon's talk of the ontic conception. He suggests, "Salmon's most penetrating idea was to abandon the idea…that explanations are arguments. Instead, he defended an *ontic* view, according to which explanations are objective features of the world" (26-27). Indeed, Salmon does at times seem to favor this purely objective take on what explanations are, as, for instance, when he expresses support for Coffa's view which, says Salmon, "located explanations in the external world" (1989, 136). Most interpreters seem to agree that this is, in fact, *the* take-home message of the ontic conception. On this way of thinking, in offering up a causal-mechanical theory of explanation Salmon aimed to specify the precise sorts of objective facts that make up explanations.

Glennan (2002), however, offers a different interpretation of 'ontic' (also see Henderson 1995, 80). Glennan claims, "[c]ausal-mechanical explanation exemplifies what Salmon calls the ontic conception of explanation. Explanations are not arguments, but are rather *descriptions* of features of a mind-independent reality——the causal structure of the world" (italics mine, 2002, 343). One finds support for this interpretation in, for instance, Salmon's complaint that Hempel sometimes slips from the epistemic proposal that explanations render happenings expectable to the proposal that they show how happenings fit into an objective "nomic nexus" (1965, 488). Salmon (1989) echoes Railton in suggesting that when Hempel slips to the latter view, "the ontic conception is being expressed" (120). Says Salmon, the ontic conception would also be expressed were Hempel to say, more aptly, that explanations show how happenings "fit into a *causal* nexus" (120). Thus, ontic theories might take many forms, so long as what they propose is that explanations (primarily) reveal something about objective states of affairs. This version of the ontic conception turns out, however, to concern the "function" or "intellectual value" of explanations (135). Explanations *are*, on this view, representations – objective facts are not in the business of revealing. Specifically, they are descriptions.

Salmon, in fact, explicitly embraces both the objective and the descriptive version of the ontic conception:

> Proponents of this conception can speak in either of two ways about the relationship between explanations and the world. First, one can say that explanations exist in the world…they are neither linguistic entities (sentences) nor abstract entities (propositions). Second, the advocate of the ontic conception can say that an explanation is something – consisting of sentences or propositions – that reports such facts. It seems to me that either way of putting the ontic conception is acceptable… (1989, 86).

Knowing this puts us in a better position to understand Salmon's pro-ontic, anti-psychologistic arguments.

2.2.2 Arguments for 'the' Ontic Conception

Like Hempel, Salmon distinguishes between two senses of 'understanding,' a psychological sense, which involve feelings of intellectual satisfaction and the overcoming of psychological uneasiness, and a scientific sense, which involves accurate representations, or knowledge (1998, 90). Salmon also offers the following ontological anti-psychologism argument, which bears a good deal of resemblance to Hempel's:

> First, we must surely require that there be some sort of objective relationship between the explanatory facts and the fact-to-be-explained. Even if a person were perfectly content

> with an explanation of the occurrence of storms in terms of falling barometric readings, we should still say that the behavior of the barometer fails objectively to explain such facts. We must, instead, appeal to meteorological conditions (1984, 13).

The central premises of this argument are roughly as follows:

> Psychologistic theories of explanation equate explanations with feelings of understanding (viz., intellectual satisfaction and overcoming uneasiness). Yet there are clear-cut cases of (objective) explanation that have nothing to do with such feelings and vice versa.

The passage then continues as follows, with what looks to be an epistemic anti-psychologism argument:

> Second, not only is there the danger that people will feel satisfied with scientifically defective explanations; there is also the risk that they will be unsatisfied with legitimate scientific explanations. A yearning for anthropomorphic explanations of all kinds of natural phenomena for example, the demand that every explanation involve conscious purposes sometimes leads people to conclude that physics doesn't really explain anything at all.... Some people have rejected explanations furnished by general relativity on the ground that they cannot visualize a curved four-dimensional space-time (1984, 13).

The idea appears to be that if we tie explanation too closely to the psychological state of intellectual satisfaction, then we risk classifying faulty explanations as good and good explanations as faulty.

Taken together, these arguments are meant to show, the passage concludes, that "[t]he psychological interpretation of scientific explanation is patently inadequate" (1984, 13). Instead, Salmon favors an ontic conception, though we should bear in mind that in this passage Salmon appears to move from the objective version of the ontic conception in the first argument to the descriptive version in the second. Explanations in the objective sense are, after all, not the sorts of things that can be bad or defective…they just are. Then again, it may be that even descriptive explanations cannot, strictly speaking, be bad either for Salmon, for he appears to adopt the Hempelian view that where accuracy is lacking, one has at best a "potential" explanation (107). Says Salmon, "I would be inclined to say that the truth of [a description] is a necessary condition for it to furnish an actual explanation of the explanandum" (108).

2.3 Craver's Anti-psychologism

One latter-day proponent of Salmon's objective-mechanical ontic conception, Carl Craver, appears to share many of these same sentiments. For instance, he claims at the outset of *Explaing the Brain*: "explanations are not developed merely for the explainer's intellectual satisfaction?[sic.]— the ineffable "a ha" feeling that comes with understanding something. Such emotions and feelings are terrible indicators of how well someone understands something…" (2007, ix). Later he claims, "The pleasure of understanding is often indistinguishable from the pleasure of misunderstanding. The sense of understanding is at best an unreliable indicator of quality and depth of an explanation" (2007, 21). This sounds a good deal like the sorts of epistemic anti-psychologism arguments considered above.

Craver favors the objective version of Salmon's ontic conception and maintains that it can, better than any psychological conception of explanation, solve philosophical puzzles (e.g., concerned with the relevance of time and causation) related to the sorting of genuine

explanations from non-explanations. At the end of his critique of Churchland's psychological prototype theory of explanation, he claims:

> A final point serves to underscore the importance of shifting attention away from the [cognitive] representations used in explanations and toward the causal structure of the world. Some phenomena might be so complex that they overwhelm our limited cognitive systems. Perhaps a mechanism has so many parts with so many interactions that it is *impossible to understand*… For this reason, neuroscientists who revel in the complexity of the brain are increasingly using computational tools and databases…that allow them to make explanatory connections that would escape them if they relied only on their unaided cognitive abilities. It would be wrong to say that the phenomena produced by such complex mechanisms have no explanation. *The explanations exist* even if we can not represent them cognitively (italics mine, 2007, 33-4).

Craver thus appears to favor an objective conception of explanation to a psychological one on (*inter alia*) the following grounds:[3]

> Psychologistic theories of explanation equate explanations with the psychological sense of understanding (which may comprise feelings of intellectual satisfaction or *aha* feelings). However, there are clear-cut instances where the (objective) explanation for a happening exists where no sense of understanding is possible.

It bears mentioning that while Craver explicitly advocates an objective conception of explanation, his main purpose in *Explaining the Brain* is to elucidate the appropriate norms of acceptance for explanations. Once again, however, a shift to epistemic considerations requires a shift to some representational conception of explanation or other. Notice also that, like Hempel, Craver maintains that an essential characteristic of such explanations is that they be accurate. For instance, Ptolemy's model of retrograde motion is, on Craver's view, no explanation at all since it delivers an inaccurate description of the objective facts (20).

2.4 Trout's Anti-phenomenalism

Another latter-day mechanist, J.D. Trout, worries that when evaluating explanations, too often "we tap into the phenomenology associated with offering the explanation, the sense of understanding conveyed by detailing it" (2007, 566). This strategy, he suggests, is epistemically pernicious, for "[t]he sense of understanding actually does harm, sometimes making us squeamish about accepting true claims that we don't personally understand, and more often… causing us to overconfidently accept false claims because they have a kind of anecdotal or theoretical charm" (578). On its face, this looks a great deal like the sorts of epistemic anti-psychologism arguments we have been considering, but Trout's target is *phenomenalism* about

---

[3] To be fair to Craver, his chief complaint against psychological models of explanation such as Churchland's prototype model, and against other representational accounts such as covering-law and unification models, is their insensitivity to the relevance of causal facts. Craver acknowledges that much psychological research has been done on how we perceive, and think, about causation (2007, 30). One might think that a theory of causation grounded in this kind of research (e.g., Waskan in press) might succeed where philosophical attempts to understand causation have failed. Craver contends, however, that we must start by determining what the actual distinction is between causal and non-causal sequences. I might agree were I not sufficiently impressed with how far the psychological research has come without the help of metaphysicians and did I not think that it is *precisely because of* our cognitive limitations that we have a concept of causation, one which I suspect largely imposes joints on nature rather than the reverse.

explanation, the attempt to link explanations with conscious feelings of understanding. Trout introduces into the discussion a body of research suggesting that pervasive reasoning biases (viz., hindsight and overconfidence) frequently bring about an irrational confidence in rectitude. At times, he even equates feelings of understanding with these feelings of confidence (571).

Trout suggests that a proper theory of explanation must leave room for the fact that some genuine explanations fail to elicit any of the phenomenology oft-associated with explanation – for instance, the feeling of understanding, feelings of insight or confidence, or *aha* episodes. Like Craver, he draws attention to hyper-complex explanations, such as explanations for speciation and disease where the complexity of the explanation overwhelms our limited cognitive abilities. However, the cases of explanation he considers are clearly descriptive rather than objective in nature (2007, 583). He suggests,

> It is always possible to reply that these *narratives* may be intellectual endeavors of some sort, but without resources of familiarity to inspire an 'Aha' experience, they are not cases of understanding. But this reply seems suspiciously ad hoc. They certainly behave like other explanations, and they are treated as such by scientists (italics mine, 584).

In other words, Trout appears to be offering something like the following ontological anti-phenomenalism argument:

> Psychologistic theories which equate explanations with feelings of understanding (or with *aha* episodes or feelings of insight or confidence) are belied by clear-cut instances of (descriptive) explanation without such feelings.

Like the others, Trout suggests that it is of the very essence of explanations that they accurately represent objective facts. Accordingly, he supports Humphreys' claim that "It is no explanation to provide a distorted representation of the world, and the 'understanding' induced by such incorrect models is illusory at best" (qtd. in Trout 2002, 213). For Trout, genuine explanation requires genuine understanding, but the latter has nothing to do with conscious phenomenology, as it may just as well result from implicit learning (2007, 581). He sums up his own "objectivist, ontic, account" of explanation as follows:

> Perhaps a suitably hedged description of conditions for *genuine explanation* will be useful. Genuine understanding exists, not when the sense of understanding dawns, but when the following objective conditions are met:
>
> 1. The explanandum-statement putatively understood is at least approximately true,
>
> 2. The agent has sufficient collateral theoretical knowledge or information (explanatorily) relevant to that explanandum, and
>
> 3. The explanandum-belief is produced by a reliable process, whether perceptual, cognitive, or social" (2007, 584).

Though he jumps here from talk of descriptions to beliefs, it seems clear that Trout maintains that explanations at least sometimes comprise psychological representations and that it is of the essence that these beliefs accurately reveal (or at least represent) objective facts. He appears, moreover, to tend towards a mechanistic view of what the relevant facts are (2002, 229; 2007, 572).

3 THE AM*TRI*GUITY OF 'EXPLANATION' AND 'UNDERSTANDING'

Though Hempel's covering-law model would be rejected by mechanists, with whom I largely sympathize, the latter appear to co-opt a number of Hempel's arguments and assumptions related to psychologistic theories of explanation. Most importantly, all parties discussed above assume that psychologistic theories of explanation at least generally equate explanations with the likes of *aha* feelings or feelings of understanding, insight, confidence, satisfaction, overcoming uneasiness, or familiarity.[4] Let us call this set of features the commonly associated phenomenology of explanation (CAPE). All of the ontological arguments attempt to dissociate, or doubly dissociate, explanation from the CAPE, and thus purport that explanation is not to be identified in any way with the CAPE. The epistemic arguments all purport that the CAPE is irrelevant to the evaluation of explanations. What, then, are we to make of these arguments? My own interests lie primarily with the ontological question of what sorts of things explanations are, so I begin with an assessment of the ontological arguments.

3.1 Ontological Arguments Reconsidered

It should be clear enough from the preceding discussion that "[t]he radical ambiguities of "explanation" and "understanding" create almost endless opportunities for obfuscation and confusion" (Salmon 1998, 9). I would contend, moreover, that all of the ontological arguments exploit the ambiguities of both terms.

To see why, notice first that there are, both in scientific and everyday contexts, at least three different ways in which we commonly use 'explanation.' Consider, for instance, the various evolutionary explanations that have been offered for the fact that humans are distinct from other primates in being relatively furless. In *The Naked Ape*, zoologist Desmond Morris suggests that the trait may have come about because our ancestors once lived along a tropical shoreline and that our furless physique enabled us to swim and gather submerged bounty such as shellfish. This would also account for other unique characteristics, such as our having a sub-cutaneous, blubber-like layer and the fact that we take to the water well, even very early in life. When discussing origins of human furlessness, one might say, "There is an aquatic-ape explanation for human furlessness on page thirty-seven of *The Naked Ape*." Here we may well be using 'explanation' to refer to a description of the possible cause of furlessness. On the other hand, if we took it to be established fact that the trait evolved as a way of cooling humans on a hot, dry savanna, we might say, "The explanation is that losing fur prevented overheating on the hot, dry savannah." Here we appear to be using 'explanation' in the objective-factual sense of the term.

There is, however, a third sense of the term that is notably absent from discussions of explanation in mainstream philosophy of science. Consider, for instance, that Morris entertains a number of further ways of making sense of furlessness, such as that it might have facilitated removal of parasites, it might have helped our ancestors to avoid disease when picking through carrion, or it might have prevented our fire-wielding ancestors from combusting. Someone else, let us call him Gould, might have knowledge of all such proposals but view them as hopelessly unconstrained by the available evidence. He might, in fact, be suspicious of the lot of them,

---

[4] Hempel talks of empathy, but I will drop this feeling from the discussion. Empathy, or something like it, may turn out to play a role in explanations for animate behaviors, particularly those of other people (see Waskan 2006). There may even be two distinct psychological systems for animate and inanimate explanations. If this is so, then it will complicate our full account of the nature of explanations for why and how happenings occur.

viewing them all as plausible-sounding, Kiplingean *just-so* stories. It seems not infelicitous to say, in that case, "Morris has a number of explanations for furlessness," or "Gould has a number of explanations for furlessness, but he has no confidence in any of them."

Consider, also, that we commonly use 'explanation' in a similar way in everyday contexts. For instance, one who has seen the movie *Castaway* might recall Tom Hank's character, Chuck Noland, and his first night on the deserted island. Noland's attempt at sleep was disturbed by thumping noises that were coming from the jungle at irregular intervals. Noland began yelling at them, and, though he never said a word, he appeared at that point to have an explanation for the noises. He believed that they were, or might have been, caused by the island's other inhabitants. The next morning, however, as Noland walked among the palm trees, he saw and heard coconuts hitting the ground. Once again, though he never said so out loud, it was clear at that point that he had come to have a very different explanation for the frightening sounds, as did we silent spectators. We often use 'explanation' in these contexts to refer to something that can be *had* without ever being described and for something that may even be quite mistaken. This last part is obviously in tension with the aforementioned assumption that accuracy is essential. I will contest this assumption momentarily, but notice for now that to have an explanation in such contexts seems to be something along the lines of having a belief about what may have produced the happening in question. There seems, in other words, to be a psychological sense of the term as well. The psychological events involved have become the focus of psychological studies of explanation.

Notice also that all of the ontological arguments we have considered attempt to establish that explanations are not to be identified with the CAPE on the grounds that there are clear-cut cases of explanation where (some or all of) the CAPE is absent and vice versa. To make these arguments as strong as possible, let us grant that in the former cases *none* of the CAPE is present and in the latter *all* of it is. Even granting this, the arguments are seriously flawed, for all of the clear-cut cases in question are either cases of descriptive explanations (e.g., an argument concerning the free-fall of some body or a hyper-complex description of speciation) or of objective explanations (e.g., the atmospheric causes of weather phenomena or the hyper-complex neurophysiological mechanisms responsible for some phenomenon). With as much force, one might argue that explanations just are psychological given that there are cases of explanation without either descriptions or accuracy (e.g., Noland's initial explanation for the noises).[5] To dissociate psychological explanations from objective ones would, admittedly, require that we first reject the assumption that psychological explanations are necessarily accurate.[6] For now suffice it to say that the problem with all of the ontological arguments is that 'explanation' is, at the very least, am*tri*guous. Because of this, we should, *modulo* the accuracy assumption, be somewhat surprised *if there were not* clear-cut cases in which one sense of the term applies but not the rest. There is, however, nothing here that should lead us to suppose that explanations just are one of the following: objective, descriptive, or psychological. Moreover, for all the ontological arguments tell us, it may yet be that explanation, in the psychological sense, really is identical with at least part of the CAPE. Before determining whether or not this is so, let us consider, at long last, whether or not explanations must be accurate.

---

[5] Some foolish individuals have, in fact, argued as much (Waskan 2006).

[6] I use 'psychological explanation' in what follows to talk about the referent of the psychological sense of 'explanation.'

Recall, again, that the assumption that accuracy is essential clearly has to do with representational conceptions of explanation (e.g., descriptive or psychological). This assumption has been widespread from inception of the philosophical debate about explanations, but it is an ill-motivated (as near as I can tell) and fairly radical departure from common usage. To see why, consider that if we, *qua* folk or scientists, commonly assumed that accuracy were essential, then we would presumably avoid using expressions such as "bad explanation," "false explanation," "faulty explanation," "defective explanation." In point of fact, even those who assume the necessity of accuracy have a hard time avoiding such expressions, but by their lights these expressions appear oxymoronic. By the same token, it would be redundant to speak of good, true, or accurate explanations, as we often do.

Perhaps more importantly, in everyday life and in science, we frequently take ourselves and others to possess multiple, competing explanations for happenings, as was illustrated above with regard to human furlessness. At the very heart of the scientific enterprise is the process of making inferences as to which of a number of often radically inconsistent explanations is best. For instance, consider the following two claims:

> The extinction of the dinosaurs was primarily precipitated by a massive, endogenous (to Earth) volcanic event in the region of the Indian subcontinent.
>
> The extinction of the dinosaurs was primarily precipitated by the impact of a massive asteroid that occurred in the Yucatan region.

It seems no violation of our normal linguistic practices to suggest that both of these are, and give voice to, genuine explanations. This, again, is not to be expected on the view that representational explanations are necessarily accurate. One might also note that even if, as some take to be the lesson of the pessimistic induction, scientist of today are radically mistaken, we would still want to say that they have succeeded in generating many perfectly genuine explanations (de Regt 2004, 107).

The claim that *we* (by which I mean layfolk and scientists alike) do not normally take accuracy to be an essential feature of representational explanations is an empirical one that must itself ultimately be adjudicated through the use of scientific methods. I have simply tried here to marshal some preliminary, *intuitive* evidence in its favor. Henceforth I will assume that accuracy is not essential (though it will be far from devastating for my view if I am mistaken). Granting this, one finds it easy to pull psychological explanation apart from the CAPE. I will revisit the other two conceptions in closing.

First let us see whether or not the CAPE is, in whole or in part, sufficient for explanation. Taking the components of the CAPE individually, the task appears quite tractable. Consider, for instance, the oft-mentioned *aha* experience. We frequently undergo it for reasons that have nothing to do with explanations. This happens, for instance, when we feel that we have discovered the solution to a math problem, a riddle, or a design problem. Terms like "Eureka!" (as commonly used) and "Aha!" appear to give voice to the feeling of having made a discovery, and they usually occur only at the moment of discovery. These feelings profoundly diminish with each subsequent pass of revisiting the discovery. [Notice also that when the same discovery is imparted to us by another, we instead undergo an *Ooooh* moment.]

It may be possible to establish, however, that even the entire CAPE is insufficient for psychological explanation. The full array of such feelings seems, for instance, to be induced at

times by neural malfunctions of both endogenous and exogenous origins. To take a semi-fictitious example, in *A Beautiful Mind*, John Nash appears to have these feelings, with great vivacity, though in the clear light of day he realizes that he has been completely delusional. He finds that the record of his supposed discovery, which would have been mathematical in nature, is meaningless gibberish. Those who take psychedelics have been known to undergo such episodes as well.

We might, on the other hand, have one or more explanations without undergoing any of these feelings at all. We may, as already mentioned, possess some radically inconsistent explanations for a happening but lack the conviction that any of them is right. In such a case, we may fail to have any *aha* feelings or feelings of understanding, insight, confidence, satisfaction, overcoming uneasiness, or familiarity. At the very least, it seems not incoherent that this could happen, but it should if any part of the CAPE were necessary for psychological explanation.

We could keep going like this, but I am happy to grant that the CAPE is neither necessary nor sufficient for psychological explanation, nor descriptive or objective. After all, were it constitutive of explanation on any sense of the term, it would presumably be constitutive of explanation in the psychological sense. Yet it appears not to be. Thus, although all of the ontological arguments considered here trade on an equivocation, there is a kernel of truth to them. One naturally wonders, then, what this tells us about psychological explanations.

Upon doubly dissociating explanation from the CAPE, some have suggested that explanations are simply not psychological. However, we do commonly use 'explanation' in a non-descriptive, non-objective, apparently psychological sense. In fact, what we have just done is doubly dissociate the CAPE *from* psychological explanations. More plausible, then, is Trout's suggestion that conscious feelings and phenomenology are not constitutive of psychological explanations. There is, however, a third option: One might argue that some form of extra-CAPE phenomenology *is* constitutive. This option has, I believe, considerable plausibility.

But what could the missing, essential phenomenological ingredient be? I think it may be understanding after all. We appear to have been improperly turned away from the consideration of understanding as a possible constituent of psychological explanation by the amtriguity of 'understanding.' There is, as already agreed, genuine understanding of how or why *simpliciter*. To have this, we must be right. Yet, rectitude, I have also argued, is not necessary for explanation on any sense of the term. We may have, and proffer, explanations that range from being mildly inaccurate to wildly so. There is, in addition, the mere *feeling* that we understand, which seems at times to occur without psychological explanation and vice versa. It seems, moreover, to require *confidence* in rectitude, but we may have explanations without this as well.

There is, however, also such a thing as understanding of how or why *possibly*. This, according to Machamer, Darden, and Craver (2000), is precisely what mechanistic, descriptive explanations supply. They claim:

> Mechanism descriptions show *how possibly, how plausibly,* or *how actually* things work. Intelligibility arises not from an explanation's correctness, but rather from an elucidative relation between the explanans (the set-up conditions and intermediate entities and activities) and the explanandum (the termination condition or the phenomenon to be explained)...." (pp. 21).

Understanding how or why possibly is, as they intimate, sometimes known as finding a happening *understandable, comprehensible* or *intelligible* (see Machamer 1998; Waskan 2006, 2008). To achieve this state, we need have neither confidence in rectitude nor rectitude itself (i.e., we need not actually understand). However, there is more to it than intelligibility than the mere *feeling* that sometimes tracks it. While the feeling may be had without explanation and vice versa, it is exceedingly difficult to find, or even imagine, a case where one genuinely has an explanation for a happening and yet does not understand how or why it *might have* occurred, and it is equally difficult to find or imagine cases where someone genuinely understands how or why possibly and yet lacks an explanation.

As a fellow mechanist, I am inclined towards the view that understanding of how or why possibly involves, at least where inanimate explanations are concerned, knowledge of how the spatial, temporal, and causal arrangement of parts might conspire to produce the happening that interests us. Much of my own research on explanation has focused on the matter of precisely what is involved in finding a happening mechanistically intelligible. The study of the psychological underpinnings for intelligibility has, I believe, potential to tell us much about what explanations are, in *all three* senses, and about the manner in which we evaluate them.

3.2 The Nature, and Epistemic Import, of Intelligibility

One of the most perplexing facts about understanding the mechanisms by which a happening may have been produced is that, in virtue of this form of understanding, we are blessed with open-ended knowledge of both the implications of those mechanisms being as we envision them and of the possible ways in which those implications might be defeated. These sorts of knowledge obviously lie at the very core of the scientific activity of testing explanatory hypotheses (Waskan 2006, 2008).

The first sort of knowledge enables us to devise, and comprehend rationale for, tests. If explanations only had implications regarding the happening in question, we would be hard pressed to find compelling reasons for preferring one to another. But explanations typically have boundless further ramifications, boundless *surplus meaning*, as it is sometimes called (MacCorquodale & Meehl 1948). The second sort enables us to devise, and comprehend, various ways of hanging onto the explanations we favor even when their apparent implications are not born out. One important way of rationalizing this form of intransigence is to determine ways in which the envisioned mechananism-implication relation might be defeated. To be sure, at the level of local rationality this form of activity can be driven by epistemically pernicious biases such as over-confidence (Trout 2007). However, at the more global level of communities of scientists, it is precisely our ability to come up with countless reasons for hanging onto our explanations that ensures that there are always many dogs in the hunt (see Lakatos 1970). Put differently, our ability to rationalize our intransigence ensures existence of multiple trajectories through accuracy space and helps science avoid getting trapped in local minima.

What I propose, in earlier work, is a unified account of what, on my view, is explicit knowledge of possible productive mechanisms (i.e., an account of intelligibility) and tacit knowledge of an explanations boundless implications and defeaters (Waskan 2006, 2008, 2010). To tie this proposal back to the discussion of Hempel with which we started, recall that Hempel abstracted not just from facts about human psychology and language use, but he also abstracted from the specific contents of descriptive explanations and attempted to discern their bare form

with the aid of techniques of inductive and deductive logic. This second abstraction itself leads, I believe, to be a local minimum as concerns the study of explanations.

As you may know, the study of explanation in philosophy began as logical positivism lost its grip, not just on physics, but on psychology. There too it became allowable to study and talk about, complicated theoretical underpinnings form human behavior. At this point, the study of explanation got underway, as did the study of human ratiocination more generally. Of particular interest in the latter respect were the many attempts, in the new field of A.I., to account for human mechanical reasoning (e.g., in the service of planning) by, in essence, letting abstract principles of formal logic shoulder the inferential load. It would not take long for researchers in A.I. to see that in order to get a system to know what we know about the implications of mechanisms, one would need to build into the system an explicit specification of not just the countless implications of alterations to a system, but of the countless ways in which each implication might be defeated. Even with regard to simple 'toy worlds,' this often turns into an intractable task. This is, in essence, the notorious frame problem of A.I. (McCarthy and Hayes 1969; also see McCarthy 1986 and Janlert 1996).[7] There is simply no way, where such an explicit specification is demanded – that is, through use of an "extrinsic" representational scheme (Palmer 1978) –for any finite system to embody this kind of boundless knowledge.

An alternative is to suppose that we humans utilize intrinsic representations. An intrinsic representation is one such that side effects of alterations to representation automatically mirror side effects of alterations to represented system (see Haugeland 1987). The paradigm of an intrinsic representation of a mechanism is a scale model. Constructing a scale model of a mechanism obviates an antecedent and explicit specification of consequences and their possible defeaters. To determine these, one simply alters the model in the relevant ways and lets the consequences play out. Indeed, about the same time Hempel began defending his logic-based model of explanation in philosophy, Craik (1952) proposed, in *The Nature of Explanation*, that humans possess explanations in virtue of carrying "a 'small-scale model' of external reality" in their heads.

An economical alternative to constructing scale models is to construct what are widely known as finite element models. Such models were pioneered in fields like mechanical engineering, but they are now used widely throughout engineering and are increasingly used by scientists. Finite element models are typically implemented using the vast, parallel number-crunching capabilities of supercomputers and, ultimately, they supply an existence-proof that brains, too, are capable of harboring and manipulating non-sentential intrinsic models (Waskan 2003). Such models, I maintain, constitute our core inference engine for determining the countless implications of mechanisms and their countless defeaters, both for prospective planning ends and retrospective explanatory ones.[8]

The resulting account of explanation (i.e., in the psychological sense), which I term the *model model*, is as follows:

---

[7] This is not to be confused with Dennett's version of the frame problem, which has more to do with time, or with Fodor's, which has to do with the holistic nature of our entire belief system. With regard to the latter, see Waskan & Bechtel (1997).

[8] It is, I would argue, through our possession of these models that we are, as de Regt puts it, able to "recognize qualitatively characteristic consequences of [an explanation] without performing exact calculations" (in press). As de Regt too recognizes that this sort of knowledge plays an important role in the evaluation of explanations.

> "to have an [inanimate] explanation is to have the belief that a certain mechanism is, or may be, responsible for producing some happening, where such beliefs are constituted by mental representations [viz., intrinsic cognitive models] of those mechanisms. It is largely in virtue of our awareness of the information conveyed by these representations that events and physical regularities are rendered intelligible" (Waskan 2008, 262).[9]

I contend, further, that these "occurrent beliefs… endow us with tacit knowledge of many kinds of evaluatively relevant information" (272) – that is, of the sort discussed above.

Regarding the ontological arguments, then, my claim is that explanations for happenings in the psychological sense are at least plausibly construed as fully dissociable from the CAPE. However, that does not mean that conscious awareness is not constitutive of such explanations. What is constitutive is that we are aware of the mechanisms that may have produced the happening in question. This occurs, I contend, through occurrent beliefs that comprise intrinsic cognitive models of mechanisms.

As for the epistemic arguments, they all suggest that the presence of the CAPE is irrelevant to the evaluation of (representational) explanations. This may be so, but there are other forms of awareness that do bear on the evaluation of explanations. The mere having of an explanation is, of course, no reason for thinking that it is the right explanation. Something that is evaluatively relevant, however, is that we are aware of the implications of our explanations and whether or not they are (i.e., directly or indirectly in terms of *their* implications) borne out by experience. If they are not, we may try to find ways in which those implications might have been defeated. This is not the whole of what matters for evaluation, but it is quite central to the evaluative process, and the proposal that it occurs through the manipulation of intrinsic cognitive models enables us to understand, for the first time, from whence this boundless evaluative information issues. Our own intrinsic non-sentential models are, on this view, the original basis for explanations, though, as discussed further below, to overcome limitations of memory we often rely upon external non-sentential intrinsic models (and sometimes extrinsic ones) as a prosthetic.

Beyond general objections to the mechanistic approach, one serious concern about the specific account of explanation offered here stems from the fact psychological explanations are here construed as a species of belief. Yet we often use belief in both an occurrent sense (i.e., where we are consciously entertaining the belief) and a dispositional sense. This likely reflects, or is some imperfect shadow of, the very real distinction between working and longer-term forms of memory. By the same token, we often take someone to have an explanation even if they are not thinking about it currently. We may, then, need to distinguish between two sub-senses of

---

[9] There are probably two distinct systems for processing inanimate and intentional explanations, though the former has historically been applied in order to understand inanimate behaviors and the latter is often applied to intentional behaviors. My focus here is on inanimate explanations, where I believe the mechanistic approach is a good fit. The inanimate-explanation system, on my view, delivers intrinsic models of how the spatial, temporal, and causal arrangements of extended parts produce the happenings that interest us. Positing intrinsic representations is, I maintain, also required with regard to explanations for intentional behaviors (Waskan 2006). One plausible solution is that we use ourselves as intrinsic models for others. To be more inclusive, perhaps I should have formulated the model model more generically as follows: To have an explanation for a happening is to have an intrinsic cognitive model of what may have produced it. To have an explanation is to have the belief that a certain process is, or may be, responsible for producing some happening, where such beliefs are constituted by intrinsic cognitive models of those processes.

'explanation' in the psychological sense. Even so, an essential feature of dispositional beliefs is their relationship to occurrent ones, though the converse does not hold. Thus, though we may have an explanation for a happening at time *t* without thinking about it at *t*, plausibly what makes it an explanation has to do with the fact that *were we to recall it* from memory we would thereby understand how or why possibly that happening occurred.

3.3 Further Grounds for Going Psychologistic

There may, moreover, be important parallels between, on the one hand, the relationship between occurrent explanations and dispositional ones and, on the other, the relationship between occurrent explanations and both descriptive and ontic ones. That is to say, these other forms of explanation may themselves be constituted in a way by occurrent beliefs about how or why possibly. As you will recall, some attempt to dissociate explanations from conscious psychological facts by pointing to cases of descriptive explanation that fail to generate feelings of understanding in some (Hempel 1965) or in all (Trout 2007). A similar point could be made regarding intelligibility. Consider, for instance, the aquatic-ape explanation on page 37 of in *The Naked Ape*. The explanation is there, in the book, even if nobody is reading, or thinking about it. Similarly, as Trout suggests, a hyper-complex model will sometimes involve so many factors that "memory could not hold them all at once" (2007, 583). In such cases, it may, at least for humans as they now are, be impossible for such models to render the happenings of which they are models intelligible to us. To the extent that we do take them to be explanations (an empirical matter), one might conclude that what makes for a descriptive explanation has nothing to do with human psychology. But that may, for various reasons, be too hasty.

It may be that the scientists who reputedly treat these descriptions as explanations only do so insofar as they have found some way to re-represent them in ways that compresses much of the complexity, albeit at the expense of important details, so as to in fact render the happenings in question intelligible. Consider, for instance, Elman's neural network models of grammar learning (REF). Like virtually everyone who works with large connectionist systems, Elman tries to understand how his work through statistical analyses (e.g., principle component analysis), many of which supply comprehensible graphical representations of their functioning. Indeed, when thinking about and discussing such models, one commonly finds invoked analogies and metaphors grounded in happenings that are comprehensible (e.g., partitioning of state space).

Even bracketing all of this, it may be that intelligibility is an essential part of what it means for a description or external model to be an explanation. It could turn out that we (*qua* folk or scientists) regard these external representations as explanations only insofar as *were one to fully comprehend* (i.e., in the linguistic sense) *the descriptions they comprise*, one would thereby find the happening intelligible. Similar considerations may apply in the case of objective explanations. It may be that we regard some set of objective facts (e.g., the sorts of hyper-complex objective facts mentioned by Craver) as an explanation for a happening only insofar as *were one to be aware of those facts*, one would thereby find the happening intelligible. If this is right, then a better grasp of what it is, psychologically speaking, to find a happening intelligible may tell us a great deal about what explanations are, in all three senses of the term discussed herein. Moreover, scarcely a philosophical discussion is to be found that does not invoke the likes of revealing, exposing, showing, enabling us to see, laying bare, elucidating, and so on (Waskan 2006; Wright and Bechtel 2007). A better understanding of intelligibility may help us to discharge these (often visuospatial and perceptual) metaphors.

If the above proposal regarding the subtle connection between intelligibility and the different conceptions of explanations is on the right track, it raises a host of further questions and suggests many avenues of further study. One wonders to whom the 'one' refers in the above analyses. Do we, *qua* folk or scientists, require that an explanation must be capable of rendering a happening intelligible to actual humans as they now exist? As we have seen, often mentioned are the quantitative limits of human memory and the limits they impose upon how much complexity we can 'wrap our heads around.' Left untouched, however, are questions of how qualitative differences in how a being conceives of the world affect our judgments. Perhaps we would count something an explanation for a happening so long as, by comprehending or becoming aware of it, that happening would be rendered intelligible to any being, even a deity, that comprehended or was aware of it. Against this view, one does find in the historical record cases where a model's *qualitative* incomprehensibility led many to deny that any explanations were on offer (Waskan 2006, 270-1). In any case, I leave these as open empirical questions to be resolved by further analysis, by culling the historical record, and by direct experimental study.

Indeed, in the promise of this last method, we find further grounds for going psychologistic, specifically with regard to the study of how and why we apply the term 'explanation' in the various ways that we do. As the above arguments show, a significant portion of the discussion of explanations is driven by judgments, whether of folk, philosophers, or scientists, about whether or not to apply the term to particular cases. Presumably the judgments we make stem in large part from our tacit command of information, gathered throughout our lives, about the socioculturally (or socio*sub*culturally) accepted standards for applying the term, from what is sometimes called a 'concept.' Empirical research on our concepts of explanation can be expected to tell us more about the nature and structure of these concepts. However, whether we study explanation through standard analytic techniques or through more scientific methods (e.g., of experimental philosophy), we are to some extent relegated to hoping that our individual or shared concepts of explanation are produced by, and somehow mirror, some more objective sets of facts (see Goldman & Pust 2002).

It may be, however, that in the case of explanation we can climb right up to the headwaters of this river and paddle our way back down. For instance, perhaps it is possible to discern the actual psychological underpinnings for beliefs about how and why possibly, the actual psychological basis for intelligibility. From here we may gain a view of the origins concept of psychological explanation gleaned from our imperfect apprehension of these psychological goings-on and revealed by philosophical and (eventually) experimental techniques. Knowing all of this will ultimately help us to better understand precisely why folk, scientists, or philosophers have the *intuitions* about cases that they do. This general sort of thinking is what led me to say, some while back, that "the Model model is a mechanistic model of the psychological underpinnings for explanation and, thereby, for our philosophical… intuitions concerning the nature of explanation" (Waskan 2006). Then my target was explanation in the psychological sense. However, if my proposal about the role played by intelligibility in the other two concepts of explanation (i.e., descriptive and objective) is on target, then this line of inquiry may help to reveal why we apply all three senses of the term 'explanation' in the precise ways that we do.

References

Craik, K.J.W. (1952). *The Nature of Explanation*. Cambridge, UK: Cambridge University Press.

Craver, C. (2007). *Explaining the Brain*. New York: Oxford University Press.

de Regt, H.W. (2004). Discussion Note: Making Sense of Understanding, *Philosophy of Science*, 71 (1), 98–109.

de Regt, H.W. (2009). Understanding and Scientific Explanation. In H.W. de Regt, S. Leonelli & K. Eigner (Eds.), *Scientific Understanding: Philosophical Perspectives*. Pittsburgh: Universitity of Pittsburgh Press, 2009.

Glennan, S. (2002). Rethinking Mechanical Explanation. *Philosophy of Science*, 69, S342–S353.

Goldman, A. and Pust, J. (2002). Philosophical Theory and Intuitional Evidence. Reprinted in A. Goldman, *Pathways to Knowledge: Public and Private*. New York: Oxford University Press, 73–94.

Haugeland, J. (1987). An Overview of the Frame Problem. In Z.W. Pylyshyn (Ed.), *Robot's Dilemma*. Norwood, NJ: Ablex Publishing Corp, 77-93.

Hempel, Carl G. (1965). *Aspects of Scientific Explanation*. New York: Free Press.

Janlert, L. (1996). The Frame Problem: Freedom or Stability? With Pictures We Can Have Both. In K.M. Ford, & Z.W. Pylyshyn (Eds.), *The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence*. Norwood, NJ: Ablex Publishing, 35-48.

Lakatos, I. (1970). Falsification and the Methodology Scientific Research Programmes. In I. Lakatos and A. Musgrave (Eds.), *Criticism and the Growth of Knowledge*. Cambridge, UK: Cambridge University Press, 91-195.

McCarthy, J. (1986). Applications of Circumscription to Formalizing Common-Sense Knowledge. *Artificial Intelligence*, 28, 86-116.

McCarthy, J., & Hayes, P.J. (1969). Some Philosophical Problems from the Standpoint of Artificial Intelligence. In B. Meltzer, & D. Michie (Eds.), *Machine Intelligence*. Edinburgh, UK: Edinburgh University Press, 463-502.

MacCorquodale, K., & Meehl, P.E. (1948). On a Distinction between Hypothetical Constructs and Intervening Variables. *Psychological Review*, 55, 95-107.

Machamer, P. (1998). Galileo, Mathematics and Mechanism. In P. Machamer (Ed.), *Cambridge Companion to Galileo*. New York: Cambridge University Press, 53-79.

Machamer, P., Darden, L., & Craver, C.F. (2000). Thinking about Mechanisms. *Philosophy of Science*, 67, 1–25.

Morris, D. (1967). *The Naked Ape*. New York: Dell Publishing Co.

Palmer, S. (1978). Fundamental Aspects of Cognitive Representation. In E. Rosch, & B. Lloyd (Eds.), *Cognition and Categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates, 259-303.

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World.* Princeton, NJ: Princeton University Press.

Salmon, W. (1989). *Four Decades of Scientific Explanation*. Minneapolis: University of Minnesota Press.

Salmon, W. (1998). *Causality and Explanation*. New York: Oxford University Press.

Trout, J.D. (2002). Scientific Explanation and the Sense of Understanding. *Philosophy of Science*, 69, 212-233.

Trout, J.D. (2007). The Psychology of Explanation, *Philosophy Compass*, 2, 564–596.

Waskan, J. (2006). *Models and Cognition*. Cambridge, MA: The MIT Press.

Waskan, J. (2008). Knowledge of Counterfactual Interventions through Cognitive Models of Mechanisms. *International Studies in Philosophy of Science*, 22, 259-275.

Waskan, J. (2010). Applications of an Implementation Story for Non-Sentential Models. In L. Magnani, W. Carnielli, and C. Pizzi (Eds.), *Model-Based Reasoning in Science and Technology*, Berlin: Springer, 463-476.

Waskan, J. (in press). Mechanistic explanation at the limit. *Synthese*.

Waskan, J. & W. Bechtel. (1997). Directions in Connectionist Research: Tractable Computations without Syntactically Structured Representations. *Metaphilosophy*, 28 (1-2), 31-62.

Wright, C. & W. Bechtel. (2007). Mechanisms and Psychological Explanation. In P. Thagard (Ed.), *Philosophy of Psychology and Cognitive Science*. New York: Elsevier, 32-79.