Scott Walden, *Review of Scott Sehon, 'Teleological Realism: Mind, Agency, Explanation'*, originally published at *Metapsychology Online*, March 2006

Commonsense psychology explains human behavior in terms of mental states such as beliefs and desires. Telemachus traveled to Sparta because he wanted to find Odysseus and believed that he could do so by talking with Menelaus, who lived there, and who was rumored to have been the last person to see Odysseus and his men. Two things seem implicit in such commonsense psychology: that the attributed mental states are real, and that the kind of explanation being invoked is causal. Thus

> [1] Telemachus traveled to Sparta because he wanted find his father

is on par with

> [2] the dam burst because of the torrential rains

both insofar as Telemachus's desire is as real as the torrential rains and insofar as the 'because' in each offers to inform the explainee about some aspect of the respective causal histories of the explanandum events.

However, the twentieth century was replete with attempts to undermine this construal of commonsense psychology. In mid-century behaviorist psychologists treated the reality of mental states with suspicion or argued that, if they do exist, logical connections between the explanans and explanandum in commonsense psychology rendered such states impotent from a causal-explanatory point of view. In later decades even cognitive scientists in good standing had their doubts. Stephen Stich argued that the mentalistic vocabulary of "folk psychology" should be replaced with a vocabulary of syntactic states that would emerge in the course of a maturing cognitive science (*From Folk Psychology to Cognitive Science*, MIT Press, 1985). Even Jerry Fodor—one of the most articulate and vigorous defenders of the causal construal of commonsense psychology—worried that "Twin Earth" concerns rendered beliefs and desires ill-suited for causal-explanatory purposes, offering in their place a new inventory of "narrow content" states (*Psychosemantics*, MIT Press, 1989).

By the beginning of the current decade such concerns seemed passé. Noam Chomsky had long since demonstrated that a robust ontology of mental states is essential to the explanation of verbal behavior ('Review of B.F. Skinner's *Verbal Behavior*', reprinted in Ned Block (ed.), *Readings in Philosophy of Psychology, Volume I*, Harvard U.P., 1980), and Donald Davidson had put to rest worries arising from logical connections between explanans and explananda ('Actions, Reasons, and Causes', reprinted in his *Essays on Actions and Events*, Oxford U.P., 1980). As the 1990s progressed Stich and Fodor began to soft-pedal their concerns about commonsense psychology and—in Fodor's case most clearly—eventually stopped endorsing their positive positions altogether. All seemed safe for the understanding of our conspecifics in terms with which Homer would be comfortable.

With the publication of Scott Sehon's *Teleological Realism*, however, we are poised to revisit this argumentative territory. This is not to say that we will repeat the arguments of the past half century. Sehon regards himself as a defender of commonsense psychology, endorsing an ontology of beliefs and desires as robust as any. But, for Sehon, commonsense psychology is not exactly what it appears to be. Beliefs and desires, while real and intimately involved in commonsense psychological explanations, are not *causes* of behavior. Instead they are states that aid the explainer in the attribution of *goals* to

agents, and it is these goals that are forefront in the explanatory endeavor.

Such is the essence of Sehon's positive program but, along the way, he undertakes much more than this. His survey and critique of the literature deals deftly with a host of central issues including: inter-theoretic reduction, supervenience, causal explanation, mental-state attribution, and even Humean motivation in relation to moral responsibility. Thanks to the clarity and organization of his presentation he accomplishes all this in fewer than 250 pages, endnotes and appendices included.

But does he convince? Yes and no. Most plausible and refreshing is his stance on the relationship between supervenience and reduction as applied to the mind-body problem. Mental properties supervene on physical properties only if there are no changes in mental facts without changes in physical facts. Most today—Sehon included—accept this supervenience thesis, at least in its "global" form that allows the physical changes to take place anywhere, including the region external to the body of the individual undergoing a change of mind. But some demand as well that the fact of global supervenience itself be accounted for by reducing mental properties to physical properties via "bridge laws" linking the former to the latter. If for every mental property there is some (or other) physical property that lawfully entails it, then (assuming these laws remain unchanged) any change in the distribution of mental-property instances will be accompanied by a change in the distribution of physical-property instances. Reduction by the discovery of bridge laws would thus account for supervenience in a very comforting way. The problem is that, notwithstanding several decades of effort well documented in Sehon's book, the requisite bridge laws have not been forthcoming. In the face of this some persist in looking for a suitable reduction, but Sehon reacts in a novel way:

> [i]n the absence of reduction, supervenience without composition indeed seems miraculous; but, because we are composed of physical particles, supervenience is no longer mysterious. In fact, it seems obvious to me that the supervenience of the mental on the physical can be explained by the fact that we are composed of physical particles (p. 131).

We have no properly documented examples of mentation in the absence of a physical substrate and the physical properties that realize our mental states will no doubt be horrendously complex. Given these two observations, it is hard to see why we should demand a reduction, or expect one to be forthcoming. All this I applaud, and I hope for a literature that takes its cue from Sehon's brave stance. It is perhaps worth adding that, even if bridge laws were discovered, it is not clear that those who demand an account of supervenience should take comfort. To say that it is a law that one property is instanced whenever another property is instanced is not to *explain* the correlation; it is, rather, merely to signal an acceptance of that correlation. Each and every law contains its own little mystery, no greater or less in magnitude than the alleged mystery of supervenience itself.

What about Sehon's positive program? Here I find myself less convinced. There is ample anecdotal and experimental evidence indicting that at a very young age humans not only understand simple causal explanations of a wide range of phenomena, but offer them as well. Indeed, as a species we are dangerously zealous in this regard, the *post hoc ergo propter hoc* fallacy being a foible of human reasoning that starts young and, unless consciously avoided, continues throughout life. It would be surprising if such a deep-seated human characteristic were not applied in the course of understanding our conspecifics. And [1] certainly *looks* similar in form to [2]. Thus the onus is on Sehon to explain both why we should not take commonsense psychology at face value and what his offered alternative construal of our day-to-day practice is. I will focus on Sehon's undertaking of the second of these tasks.

Explanations are what are offered in answer to certain types of 'why' questions. More specifically, explanations have the aim of *informing* the explainee with regard to some aspect of the world about which he or she is ignorant. In instances of *causal* explanation it is some aspect of the causal history of the explanandum event about which the explainee is ignorant. In order to inform successfully, the explainer must size up the epistemic status of the explainee, compare this with what the explainer knows about the causal history of the explanandum event, and then utter a sentence—paradigmatically a 'because' sentence—that relieves the explainee's ignorance. Thus a successful causal explanation must be both true (the explainer must mention an event or state that really was part of the causal history of the explanandum event) and appropriate (the event mentioned must be one the explainee is not already aware of, is one of a sort in which he or she would be interested, etc.). Explanations thus have both truth-conditional and pragmatic components. In instances of *psychological* causal explanation, it is a belief or a desire of the agent about which the explainee is ignorant, and it is the task of the explainer to figure out the relevant gap in the explainee's epistemic status and plug it. In the case of [1], the explainee is ignorant of some aspect of Telemachus's psychological constitution prior to the journey to Sparta. The explainee wants to know which of many possible desires was in fact possessed by Telemachus and causally operative in his psychology. Was it a desire for a good feast? Was he nostalgic for the company of Menelaus? No, it was a desire to find his father. To mention this in [1] is to say something that is both true and appropriate given the epistemic status of the (imagined, typical) explainee and, in this way, it successfully informs. [1] expresses a good explanation.

What about teleological explanation? For Sehon it is the agent's *goals* about which the explainee needs to be informed. [1] can be recast as

[1'] Telemachus traveled to Sparta in order to find his father

suggesting that the state of affairs in which Telemachus finds his father is the goal towards which his behavior was directed. The explainee is ignorant of this goal and, in being presented with [1'], has his or her ignorance relieved and thereby comes to understand why Telemachus undertook the journey.

However, for Sehon psychological explanation is typically not a situation in which a knowing explainer informs an ignorant explainee regarding the operative goal. Instead, all parties to the instance of explanation are ignorant, and must therefore apply *interpretive* measures in order to determine which goal was operative. More generally, in the course of deriving a teleological explanation of the form

[3] the agent φd in order to ψ

explainer-explainees apply two interpretive principles:

[I$_1$] find a ψ such that φing is optimally appropriate for ψing, given a viable theory of the agent's intentional states and circumstances

and

[I$_2$] find a ψ such that ψing is the most valuable state of affairs towards which φing could be directed, given a viable theory of the agent's intentional states and circumstances (pp. 146-147).

In applying [I₁] to Telemachus, the explainer-explainee uses his or her knowledge of Telemachus's beliefs (he believed, for example, that Menelaus was the last person to have seen Odysseus), desires (he desires to find his father), and circumstances (the situation at the palace is dire and only the strength that would come with Odysseus's return will bring affairs to proper order) to determine the goal towards which the journey was directed. There may be other candidate goals related to finding his father—an opportunity for time away from the tired palace routines, an occasion for a feast absent the suitors, a chance to drive an opulent chariot—but these are ruled out by an application of [I₂], which requires that the goal with the greatest value to the agent be designated as the operative one in the course of explanation. All these interpretive factors together enable the explainer-explainee to triangulate uniquely on the operative goal, thus relieving his or her ignorance and fostering understanding.

There are at least two problems with this approach. First, one may worry about the ontological status of goals and goal-directedness. With regard to the former, Sehon gives us a great deal of information about how an explainer-explainee *finds out* what goals an agent had—a process that involves [I₁], [I₂], knowledge of the agent's desires and a host of assumptions about the agent's rationality—but he is coy about the nature of goals themselves. At places he says goals are future states of affairs but the ontological status of a future state of affairs is itself a mystery. Perhaps Sehon believes there is no more to having a goal than what gets ascribed by the ideal explainer-explainee to an ideal agent in an ideal application of the interpretive apparatus. But, by my lights, this would mix poorly with the idea that he is offering a realist form of explanation as advertised in the title.

Sehon is much more explicit about the ontological status of goal-directness stating that he doubts that "there is *any* reductive analysis of *goal direction*, just as there is arguably no reductive analysis of what it means to say that one event *caused* another" (p. 137). I am sympathetic with the suggestion that causation will have to remain an unanalyzed primitive but to declare a stand-off on this basis is methodologically suspect. Causation is an indispensable component of our understanding of our world (as Sehon would readily acknowledge) and we tolerate its unanalyzability because we have no choice. But newly offered metaphysical notions receive no such preferential treatment, especially ones that arguably are in tension with our larger, causal, understanding of the world.

A second problem with Sehon's positive program is the suspicion that teleological explanation is merely causal explanation in disguise. The ease with which we can switch between sentences such as [1] and [1'] suggests this, as does the close connection between desires and goals. As Sehon acknowledges, desires are a rich source of information about goals. We know that Telemachus had the goal of finding his father once we know that he had the desire to find his father. But, given this, it is arguable that all the explanatory work is done once we have learned what the *desire* is and the further switch from 'Telemachus wanted to find his father' to 'Telemachus had the goal of finding his father' amounts to a mere pleonasm and thus is not a reliable guide to ontological postulation.

Sehon is well aware of this danger and provides a complex argument in an attempt to ward it off. Contained within the general form of a teleological explanation offered at [3], he suggests, are two counterfactuals which together constitute the truth-conditional core of teleological explanation:

> [4] ceteris paribus, if [the agent] A had not had the goal of ψing, then A would not have φd

and

[5] ceteris paribus, if ψing had required πing [instead of φing, or in addition to φing], A would have πd [instead of, or in addition to, φing] (p. 159).

Applying this to Telemachus yields:

[4'] ceteris paribus, if Telemachus had not had the goal of finding his father, then he would not have traveled to Sparta,

and

[5'] ceteris paribus, if finding his father had required traveling to Troy instead of to Sparta, or in addition to Sparta, Telemachus would have traveled to Troy instead of to Sparta, or in addition to Sparta.

According to Sehon, both must be true (on a reasonable construal of the truth conditions for counterfactuals) in order for [1'] to be a correct explanation.

Causal explanation also has a counterfactual truth-conditional core but, according to Sehon, it involves only

[6] ceteris paribus, if A had not had the *desire* to ψ, then A would not have φd

thus yielding

[6'] ceteris paribus, if Telemachus had not had wanted to find his father, he would not have traveled to Sparta.

[4'] and [6'] are very similar, Sehon grants, but the two ways of explaining remain distinct due to the absence of anything along the lines of [5'] in the truth conditions for [1].

I am sympathetic with Sehon's analysis of [1] in terms of [6] and [6']. If I ask you why Telemachus went to Sparta and you utter [1] and yet [6'] is false, you have misinformed me as to the cause of his journey. Similarly, the analysis of [3] in terms of [4] seems plausible. If, in answer to the same question, you utter [1'] even though [4'] is false, something seems amiss.

But I wonder whether there is in fact anything along the lines of [5] built into [3]. According to Sehon, [5] has— roughly—the force of saying that "A would have done whatever it took to ψ" (p. 159). This seems odd. The general idea, recall, is that in an instance of explanation the explainee is ignorant of some state or event relevant to the explanandum event. In the case of causal explanation the explainee is ignorant of an event that was part of the causal history of the explanandum event and, hence, the plausibility of [6]. In the case of teleological explanation, the explainee is ignorant of the goal towards which the explanandum event was directed and, hence, the plausibility of [4]. But mentioning a particular goal in the course of explaining a particular behavioral event when that goal is compatible with *any* sort of behavior seems uninformative. In the case of Telemachus we want to know why he went to *Sparta*. It is no help to be informed that he would have gone *anywhere*. Thus I am not convinced that [5] is in fact part of an analysis of [3] and yet, by Sehon's own admission, it is [5] that is warding off the charge that teleological explanations are causal explanations in disguise.

So, as I see it, there are problems with Sehon's positive program. However, perhaps I am being unfair in ignoring Sehon's argument against the causal construal of commonsense explanation. Sehon's overall rhetorical strategy is to highlight problems with the causal construal, characterize his teleological replacement, and then claim that problems with the former are more formidable than problems with the latter. Perhaps the problems I have indicated here are dwarfed by the problems Sehon finds with the traditional causal construal of commonsense psychology. I leave it to the reader to evaluate this aspect of Sehon's program and judge for him- or herself which is the less problematic approach. At the very least, however, in its style and willingness to challenge pervasive dogmas, Sehon's book is precisely what we need to help us further refine our understanding of explanation, especially explanation as manifested in psychological contexts.