Common sense psychology and scientific explanation

Frederick Stoutland

Department of Philosophy, Uppsala University

Fred.Stoutland@filosofi.uu.se

ABSTRACT: In his recent book, *Teleological Realism*, Scott Sehon defends a teleological account of explanations in common sense psychology [CSP], arguing that if such explanations were causal, CSP would be reducible to physical science. He asserts that since it is not thus reducible, its success in explaining human behavior is a mystery. I contend that many CSP explanations are causal, although in a different sense than the causal explanations of physical science. I set out the distinctive features of CSP, object to the physicalist claim that explanations in physical science are *the* basic type, and argue that CSP explanations do not need external support from physical science and that reflection on how they work dispels any mystery about their success.

I

In a recent book,[1] Scott Sehon defends a teleological account of explanations in common sense psychology[2] [henceforth: "CSP"] against the current orthodoxy that such explanations of behavior are causal. He has a particular interest in the relationship of explanations in CSP to explanations in physical science. There are, he argues, three options: 1) CSP and physical science contradict each other; 2) physical science entails CSP or vice versa, or they entail each other; 3) CSP and physical science are logically independent. The first two he regards as versions of "strong naturalism" and rejects in favor of the third option, which he accepts, although reluctantly, because it entails that the success of CSP is not explained by physical science. The latter, as Sehon admits, assumes that *any* explanation of why CSP is successful would come from physical science (including neuroscience and cognitive science), and since it yields no such explanation, Sehon concludes that we have to accept the success of CSP as a kind of mystery.

> …In the end the most basic teleological facts and concepts are irreducible and primitive. In and of itself this is not so embarrassing, for all theories… will leave some basic facts unexplained. However, a theory of the world that managed to subsume [the basic principles of CSP] by showing they followed from more basic physical science would, all things being equal, be superior to one that left [those principles] as brute, irreducible facts.

[1] *Teleological Realism: Mind, Agency and Explanation* (MIT Press, 2005). References to this book will be put in brackets in the text.

[2] This is a good enough term for the constitutive understanding that underlies human interactions with each other and the world. It is much better than "folk psychology", which wrongly suggests that this understanding is some kind of proto science that might be undermined by scientific theories. The term "common sense psychology" can be misleading if it suggests that such understanding consists of explicit knowledge or that its claims have the foundationalist status "commonsense philosophers" have given it. I take it to be a kind of practical know-how required for competent human agency whose propositional commitments can be made explicit only by hard work.

Printed from: *Hommage à Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz.* Ed. T. Rønnow-Rasmussen, B. Petersson, J. Josefsson & D. Egonsson, 2007. www.fil.lu.se/hommageawlodek

Since we cannot show that the principles of CSP follow from physical science, their holding "reliably of human beings and other agents is left as a brute, irreducible fact with no further explanation." [219]

Sehon's discussion of CSP and its relation to physical science is illuminating, and I agree with many of his criticisms of various views, particularly of physicalist attempts either to reduce CSP to physical science or to show that supervenience establishes explanatory links between them. I do not, agree, however, with a number of his central claims, and in what follows I shall try to show how I think they go wrong. I do not do this because I think Sehon has written a bad book but precisely because he has written a good one, which nicely formulates the issues so as to invite useful criticism and has enabled me to articulate my own views more clearly and coherently.

I shall focus my discussion on three claims that Sehon defends at some length.

1. The explanations of behavior in CSP are teleological; causal explanations belong to the domain of physical science.

2. Causal explanations in physical science are the fundamental kind, not least because it is intelligible to think they could explain the success of CSP explanations but not the other way around.

3. The success of CSP is, nevertheless, unexplained because only physical science could explain it and only if CSP were reducible to it, which it is not.

I agree that CSP is not reducible to physical science but I reject the other claims and hold instead the following:

1. Explanations of behavior in CSP are often causal.

2. Causal explanations in physical science are not *the* fundamental kind of explanation: what is fundamental depends on the context.

3) CSP explanations do not need and cannot have an external foundation. Their success is no mystery. for it is intelligible apart from external support.

II

The first claim is two-fold: explanations in CSP are essentially teleological and causal explanations belong to the domain of physical science. The heart of Sehon's book is the former, which deserves considerable discussion, but it is not my main interest here so I will be somewhat quick.

Sehon characterizes a teleological explanation as explaining an agent's behavior "by citing the *purpose* or *goal* of the behavior in question…" Such explanations "cite a future state of affairs toward which the behavior was directed, rather than an antecedent state that caused the behavior." [13] Consider, for example, Joan's going to the kitchen in order to get a glass of wine. Although we might express that by saying "Joan went to the kitchen because she wanted to get a glass of wine," its form in "explicitly teleological language" is "Joan went to the kitchen in order to bring it about that her desire for wine was satisfied." The "paradigmatic form" of a teleological explanation, he writes, is "A φd in order to ψ." [149]

Underlying such explanations are two principles whereby "the agent is rational and her behavior makes sense" [139] that must apply when an agent's behavior is irreducibly teleological:

> R-1: Agents act in ways that are appropriate for achieving their goals, given the agent's circumstances, epistemic situation, and intentional states.
>
> R-2: Agents have goals that are of value, given the agent's circumstances, epistemic situation, and intentional states. [139]

These principles enable us to determine whether an agent is really directing her behavior and, if so, to what state of affairs, enabling us to rule out inanimate objects, plants, and various kinds of animals as engaging in behavior that requires a teleological explanation. Although we can characterize the behavior of the latter in what appear to be teleological ways ("A heat-seeking missile turns toward the north," "A plant turns toward the sun"), "we would want to resist any attribution of agency" [161] to them because in such cases teleological explanation can be reduced to causal explanation. The two principles do not apply, for example, to the wasp, whose behavior can be adequately explained in terms of its evolutionary niche.[162] It is highly unlikely that "the wasp has a system of values and directs its behavior… to achieving appropriate goals [and hence it] is not an agent directing its behavior." Cat behavior, by contrast, "seems sufficiently rich and sophisticated to warrant attribution of a relatively complex set of goals. Moreover, we can get a grip on the idea that various kinds of states of affairs are of value from the cat's perspective: eating, being fed, being let outside…etc." [166]

Human beings are, of course, the paradigm of rational agents, and the two principles apply fully to them. The principles imply that a wide variety of counterfactuals hold of the agent, and this enables us to distinguish between an agent's merely having a goal and her having a goal *and* acting in order to achieve it. We can establish that A φd in order to achieve X rather than Y by asking such questions as "Would A still have φd if circumstances were such that φing would achieve X but would be detrimental to achieving Y?" We look, that is, "at counterfactual situations to see what account of the agent's behavior makes the most rational sense."[158] Thus Sehon counters Davidson's defense of a causal account of action explanation by appealing to counterfactuals that in his view are grounded on teleology and are not causal.

There are two objections I shall make to this teleological account of CSP explanations. The first is that it assumes that all CSP explanations of intentional action cite a goal toward which the action is a means. In my view, this is not the case: what is essential to CSP explanations is that they cite the reason for which an agent acted – they are *rational* explanations – not that they cite the purpose or goal of the action.

We typically make what Sehon calls "rational sense" of an agent's behavior by determining the reason for which she acted intentionally as she did, and that *may* require explaining her action as her taking a means toward an end. For example, we may explain why Joan went to the kitchen by saying she did so in order to get a glass of wine, which is the reason for which she acted. But rational explanations are often not like that. I wave at a friend across the street not as a means to an end, not to fulfill some goal, but simply as a friendly act. My behavior was not directed at a future state of affairs but was a response to a present one: the reason for which I waved was that my friend waved at me. I comfort a child who has fallen off her bicycle not as a means to an end but simply as someone who needs comforting and that I am in a position to provide. My swimming in the morning is not directed at a future state of affairs; I swim just because I enjoy it. Although one might be able to reformulate these rational explanations in means-ends terms, doing so would distort them because we often act intentionally without aiming to fulfill a goal.

The second objection is that Sehon confounds a teleological *description* of an action with a teleological *explanation* of it. Explanations in general depend on how the explanandum is described, a point crucial to CSP because giving an adequate description of behavior can be as difficult as explaining it. A description of behavior that permits a rational explanation of it is invariably teleological: it describes what the agent intended or meant by his behavior. When I waved at a friend, I moved my body and limbs in many ways that could have been my doing diverse things, but the crucial thing is what I *intended by* the movements I made – what I *intended* in moving my body and limbs.

To specify what I intend is to give a teleological *description* of my behavior – not to explain it but to describe it so it can be explained. The teleological *description* is that I was waving: by moving my body and limbs I intended to be waving. The explanation is that I did so in response to my friend's waving at me, which is not a teleological *explanation.* At another time when I intended my behavior to be waving, the explanation itself might be teleological in form: I was waving in order to get a taxi. The general point is that a rational explanation requires a teleological description but not vice versa.

Sehon writes that "if we want to make rational sense of an action, we want to know what the agent was trying to accomplish." [177] That is true but what is thus characterized is not the *explanation* of an action but its teleological *description* as what the agent intended. An explanation specifies an additional point, namely, *why* she acted as she did, *why* she was trying to accomplish whatever it was. This is not necessarily something we only know *after* we know what the agent intended to do: making rational sense of an action requires knowing *both* what she was intending to do and why she was doing it, and we often do not know either without knowing both.

Sehon's account of reasons for action conforms largely to the belief-desire model except that he denies that beliefs and desires play a *causal* role. While I think that model of reasons for action is deeply mistaken,[3] I won't discuss it here except to note that Sehon's treatment of Dancy's alternative to the model also confounds description with explanation. Dancy's view (which I largely share) is that reasons for action are not psychological states but states of affairs in the world to which agents respond. Sehon discusses Dancy's example of a man whose action is explained in terms of someone's having lied to him, which Dancy formulates as "The ground on which he acted was that she had lied to him." Sehon comments that "citing a previous state of affairs in the light of which the agent acted doesn't by itself make rational sense of the action…. That she lied to him may be one part of the story behind what the agent is doing, but doesn't yet tell us what the agent is trying to accomplish." [177] But Dancy's citing the fact that she had lied as the reason for the man's action *assumes* that he knows what the agent was trying to accomplish; knowing the latter *is* knowing what he did intentionally. Knowing that is, of course, part of what it is to make rational sense of the man's action but it doesn't *explain* his action – it doesn't specify the reason for which he acted. Even if knowing the reason for which he acted were necessary to know what he was trying to accomplish, it does not follow that the latter *explains* why he acted.[4]

[3] See my "Responsive Action and the Belief-Desire Model", *Graz Philosophischen Studien*, 61 (2001), pp. 83-106.

[4] Sehon's view of reasons is somewhat obscure. The clearest statement of it is this: "…If the agent φd in order to ψ, then we might say that the agent's reason for φing was whatever made ψ valuable from the agent's perspective. In other words, the agent's reason will be whatever explains the value of ψ." He goes on to argue that, although a desire can be a reason in this sense, it often is not. I think this puts reasons on the wrong level. To say that S φd in order to ψ is to say that the reason S φd was that φing was a means to ψing. But we can go on to ask why he took that to be a reason –

III

I come now to Sehon's claim that causal explanations belong essentially to the domain of physical science, which he takes to be a major reason for regarding explanations in CSP as teleological and not causal. I agree with his general way of distinguishing between explanations in CSP and in physical science. He maintains that CSP explanations are *normative* in that they involve considerations we take to favor (or disfavor) our acting, which is not the case in the physical sciences where phenomena are never explained (or described) in terms that favor (or disfavor) their occurrence – that take them to be appropriate, justified, or correct (or the reverse). CSP explanations are also normative in that ascribing propositional attitudes to an agent "involves an irreducibly normative element" [62] in the sense Davidson spelled out.[5] Davidson noted further that this implies that there are no strict law-like generalizations in CSP connecting intentional actions with psychological states since ascribing such actions and states presumes that the agent is rational, which entails that any generalizations we formulate will have to be given up if required to preserve the rationality of the agent – something we could not do if the generalizations were strict and law-like. CSP explanations do not involve precise generalizations (not even probabilistic ones) but at most rough generalizations about actions, reasons, and circumstances.

That CSP explanations are normative means they are also *agent-centered* in that they require identifying both what the agent took herself to have done and what she took to be a reason for her to have done it. An agent cannot always articulate what she did, for "taking" denotes a stance that may be implicit, and others may (in certain situations) be better able to articulate what she did (or her reason for doing it) than she can. What others articulate should, nevertheless, be an articulation of what the action was from the agent's point of view. CSP explanations can, therefore, also be characterized as *first-person*, which does not mean the agent has introspective or infallible knowledge of her own actions and reasons or that the latter cannot be known by others, but that the agent's point of view on the world is central to a CSP explanation.

The reason for this is that the normative significance of states of affairs for an agent – their practical significance as reasons that might explain her actions – manifests itself only to those who are able to view them from that agent's point of view. This is not the case for physical science, which, whether it requires a "view from nowhere," certainly does not require the agent's own point of view since its explanations may be quite unintelligible to the agent herself. To understand what an agent responds to as a reason for her action, however, requires grasping the agent's own point of view, not to agree with it but to understand how the world and its normative significance would appear from that point of view.

This implies that CSP explanations are *interpretive.* Their ascription is holistic in that our criteria for establishing the reason that explains why an agent acted is not independent of establishing what she did, and neither is independent of what she took to be

not only why he took it to be true that φing was a means to ψing but why he took that to be a reason for him to φ, which may very well be that he took ψing to be valuable (with there being a further explanation of that). The reason he took ψing to be a reason for him to φ is not, however, itself a reason for him to φ.

[5] This does not mean that a CSP explanation must justify an action or provide a reason that really favors it; it means rather that it uses terms (appeals to considerations) that could justify the action or be a reason for it.

a reason for her to do it, which may require establishing what she believed or desired, which may require finding out what she did, and so on. We cannot, therefore, simply use our own concepts and distinctions in describing and explaining an agent's action but must *interpret* how she understands it. This supports Sehon's claim that "the behavioral evidence available to the interpreter will generally allow for indefinitely many attributable states of beliefs and desires" so that she must "choose the set that maximizes the rationality of the agent."[60]

A final feature distinguishing CSP explanations is that they are *constitutive* in that a human agent who is capable of acting intentionally must also be capable of giving and receiving reasons for the actions of himself and others.[6] Such explanations are not in the domain of experts but are essential to human thought and action, a point that figures in Sehon's rejection of eliminativist views of CSP.

Sehon says rather little about the nature of the physical sciences other than that CSP is not one of them. In my view, he underestimates the complexity of the physical sciences and the unlikeness of being able to reduce either biology or chemistry to physics. He is too quick, therefore, to accept the completeness of physics, is too speculative about "the propositions ultimately put forward by a completed physical science"[9], and is mistaken in claiming, for instance, that "biological explanation no longer involves anything above and beyond the mechanistic principles of physical science." I will return to this, although it does not bear on his general characterization of physical science, which he takes to be centered on explanations of phenomena in terms of natural laws[7] that often enable causal explanations of phenomena, in particular of the behavior of organisms, including human beings. With all of this I agree.

But I do not agree that causal explanations play no role in CSP. Sehon gives no analysis of the concepts of causality and explanation, and although he does not discuss whether we should distinguish "cause" from "causally explain," he assumes that "cause" means "causally explain" and that the latter involves causal laws, which he takes to be law-like, universal generalizations in the standard sense. This is the concept of causality that underlies the notion that causal claims are founded on our ability to intervene in nature to bring about an effect, the point of such intervention (or manipulation) being to verify a necessary connection between cause and effect.[8]

I agree that explanations in CSP are not causal in *that* sense: they are not grounded in law-like universal generalizations, which is what philosophers (including my past self) usually mean when they deny that rational explanations of action are causal. I now think, however, that to confine the term "causal explanation " to that sense is to invite serious misunderstanding and that we should take it as only a species of a more general notion of causal explanation.

[6] This is not inconsistent with the claim that there are autistic persons who suffer from "mindblindness" – typically understood as an impaired capacity to ascribe mental states to others. An *impaired* capacity does not mean *no* capacity, and the impairment is much more a matter of articulating various claims about other persons than about interacting with them in linguistic and non-linguistic ways. The latter is what is constitutive about CSP explanations, which, as noted above, may be implicit without being articulated.

[7] What a natural law is and how it functions in explanation is a controversial question. The controversy has been renewed by Nancy Cartwright's claim that, taken in the traditional sense, the laws of physics are false, which requires reconstruing them in a quasi-Aristotelian sense as natural capacities.

[8] This is von Wright's view of causality as developed, for instance, in his *Explanation and Understanding*.

What is this general notion of causal explanation? I would argue that the generic sense of "explain" is "render intelligible" and that there are numerous ways to render phenomena intelligible. One might specify their parts or the whole of which they are parts, spell out their function or articulate the role they play in a narrative, clarify what to think or say about them, perhaps by analyzing terms, elucidating various claims, or redescribing them. One might show that the puzzle or mystery that motivated the search for an explanation was not really puzzling or mysterious, or trace out what claims about them imply or follow from, or what it would be for them to be true. Or one might construct a causal explanation of the phenomena.

What, then, are is required for an explanation to be *causal*? The dictionary defines "cause" as "something that produces an effect, result, or consequence."[9] Anscombe formulates the basic idea as follows: "Causality consists in the derivativeness of an effect from its causes… Effects derive from, arise out of, come of, their causes."[10] The essential thing, I suggest, is that causal explanation makes it intelligible *why* a phenomenon begins to be when it would otherwise not be, ceases to be when it would otherwise continue to be, or continues to be when it would otherwise cease to be. Call effects of that kind *events* or *processes*; what causally explains such effects *produces* them. If a cause produces an effect – if the effect *derives from, arises out of,* the cause – it's not mere luck that there is an effect: the effect would not have been had the cause not been. The latter makes the account causally *explanatory,* provided it describes the cause in a way that makes it intelligible why the effect came to be.[11]

This characterization allows for different types of causal explanation. One type is based on law-like, universal generalizations, but another type is consistent with the features that are distinctive to CSP. The latter is exemplified by rational explanations that are best characterized as causal because their effects, intentional actions, are events or processes, and because citing the reason explains why the agent acted. It not only describes what he did or what his intention or goal was, but it explains *why* he did it: what his acting derived from or arose out of, what produced it. Explaining why someone opened a door, stopped the car, or continued to climb the hill by citing the reason for which he acted says what each action derived from – what produced it.

Sehon denies that explanations in CSP are productive: "When we give a genuinely teleological explanation of a piece of behavior,… we are seeking to know the state of affairs toward which the agent's behavior was directed," [153] but we do not thereby "gesture at some sort of mysterious noncausal means of bringing about behavior." On his view, behavior has causes that bring it about, but they are *physicalistic* "with brain states playing a central role. Teleological explanations simply do not purport to be identifying the cause of a behavior."[12] [218] To explain, for example, why someone is walking to the cooler, we cite the purpose of her action – to get a beer – which is the goal "toward which her behavior was directed." But that CSP explanation does not tell us what produced or causally explained her behavior; for the latter "we can surmise that a sensory stimulus triggered a

[9] *American Heritage Dictionary* (Houghton Mifflin, 1982)
[10] *Causality and Determinism* (Columbia University Press, 1974), 136.
[11] I agree with Davidson that causes and effects can be described in purely extensional ways and hence in explanatorily empty ways, but I do not find his distinction between "cause" and "causally explain" to be helpful in an account of CSP explanations.
[12] Cf. 203: "…If mental states can causally explain behavior, then mental states must be brain states."

chain of events in her brain and nervous system, with the ultimate result that she walked to the cooler." [137][13]

On my view, although we should not rule out physicalistic causal explanations of the behavior involved in our intentional action, that does not rule out causal explanations from CSP.[14] The latter typically appeal to reasons for action, that is, to considerations that *favor* an agent's acting in a certain way. Of course, rational explanations frequently appeal to considerations that do not *actually* favor the action. Klaus gave money to a beggar because she is poor, but she may in fact be rich, her begging being a kind of theatre. Although the reason he gave the money did not actually favor his doing so, it nevertheless explained it, something we come to know by understanding his point of view – how the situation presented itself to him, namely as a poor woman who needed money.

I contend this is a causal explanation of why Klaus acted because it goes beyond saying what he did, what goal he had, or the intention with which he acted, to explaining why he acted – what his acting derived from. It explains his acting by specifying the reason without which there would have been, on this occasion, no such acting, which is to explain it *causally*. But the explanation presumes no generalizations, not even such a cautious one as, "Whenever persons of a certain type encounter, in these circumstances, a beggar they take to be in need, they give her money." We may be able to *predict* that Klaus would give a beggar money because we've seen him doing so many times, but that rough generalization does not *explain* why he does so, certainly not why he gave money to *this* beggar. What explains that is that she, as he saw it, needed money. That is the reason for his having given – its cause, but not its necessitating cause.

Anscombe's arguments that causes need be neither necessitating nor nomological I find convincing. If an event occurs that was not necessitated by law, it does not follow that it occurred by chance; its occurrence may be explained and hence not a matter of chance, and yet not a matter of necessity. That C causally explained E on this occasion does not entail that when C occurs again, E must, as a matter of law, also occur – even if the situations are the same. The claim that C was sufficient for E means that C was *enough* to bring about E – that it was a sufficient condition in *that* sense – but it need not be a sufficient condition in the logico-philosophical sense of *whenever* C occurs, then E occurs. 'Sufficient condition', Anscombe wrote, "sounds like 'enough', and one certainly *can* ask: 'May there not be *enough* to have made something happen – and yet it not have

[13] I don't think Sehon is consistent in arguing that teleological explanations, as he understands them, are not productive. He argues [159] that, although teleological claims support counterfactuals, they support different counterfactuals than causal claims do. But his example compares 1) "A φd in order to ψ", the teleological claim, with 2) "A's desire for ψ caused her to φ," which he takes as the *causal* claim. 1) supports 3) "Ceteris paribus, if A had not had the goal of ψ ing, A would not have φd", while 2) supports 4) "Ceteris paribus, if A had not desired ψ, then A would not have φd" which, he concludes, shows that what the *teleological* claim supports is different from what the *causal* claim supports. But, I would argue, the reason 1) and 2) support different counterfactuals is because *in-order-to* claims are not equivalent to *desire* claims, not because 1) is teleological and 2) is causal. *Both* are *causal* in making claims about what would not have happened had not such and such not happened, which are surely productive claims. Sehon here treats "A φd in order to ψ" as not merely stating the goal of A's action but as stating that A would not have φd if he had not had the goal to φ, which is surely a productive claim about A's φing. Further evidence of this is his discussion of counterfactuals in CSP when he writes of our knowing "what Joan would have done had she *believed* that there was no wine in the kitchen."[225] His claim that this is teleological and not causal is not credible.

[14] I deal with objections to this compatibility claim below.

happened?'"[15] Rational explanations are like that: the presence of a beggar was, on this occasion, sufficient (enough) reason for Klaus to have given her money – it explained why he did so – but it doesn't follow that if he were to encounter a beggar again, he would give her money, even if the circumstances were the same.[16]

Davidson is credited with having resurrected from the Wittgensteinian tomb the claim that reasons are causes, and his "Actions, Reasons, and Causes" was undoubtedly immensely influential in making the standard causal story the orthodox view. While I continue to be highly critical of that standard story, it is important to note that it is not in the spirit of Davidson's own account of rational explanation, which is heavily indebted to Anscombe's *Intention* and hence to Wittgenstein's observations on action. Most versions of the standard story ignore Davidson's claim that rational explanations are not reducible to explanations in physical science but belong to CSP, which he takes to have the distinctive features I sketched out above. Many overlook his claim that causal *relations* obtain *only* between *events*, and since beliefs and desires are not events, they are not causally *related* to actions. His view is that desires and beliefs causally *explain* actions but are not instances of causal *relations* connecting desires and beliefs with actions. The real explanatory force of a rational explanation, on his view, turns on the conceptual and normative principles implicit in our *interpretation* of the actions of rational agents in the light of their reasons.[17]

When Davidson is construed in this way, his view is not so very different from the one I would defend . We agree that rational explanations of action are causal but not in the sense in which physical explanations are. Where we differ is, first, in the role of beliefs and desires. I agree that they are reasons for action only in virtue of their content, but I think that *only* their content – worldly states of affairs – constitutes them reasons for action and that such content is not confined to beliefs and desires but is that to which agents can respond directly.[18] Secondly, I disagree with his claim that when a reason causally explains an action, there must, first, be an event *associated* with the reason that is causally *related* to the action and, second, there must be a law that connects a *physical* description of the *associated* event with a *physical* description of the action – hence a law of physics. While Davidson denied that laws of physics play a role in CSP itself, he held that they underlie CSP explanations: the latter *entail* the existence of correlated physical laws of which we may have no knowledge.

Although I do not accept this audacious claim, I will not spell out my objections here. I do not think it commits him to epiphenomenalism since that would require that a rational explanation is valid only *in virtue of* such a physical law, which Davidson denies in holding only that a rational explanation *entails* that there is such a law.[19] Nor do I think it supports Sehon's assumption that to account for the success of CSP we must appeal to physical science, a point to which I return below.

[15] G. E. M. Anscombe, "Causality and Determination" in *Metaphysics and Philosophy of Mind* (University of Minnesota Press, 1981), p. 135.

[16] Kant wrote in the *First Critique* [A549, B577]: "Every cause presupposes a rule according to which certain appearances follow as effects; and every rule requires uniformity in the effects. This uniformity is, indeed, that upon which the concept of cause (as a faculty) is based…." I accept this if it means same cause-same effect: if cause C brings about effect E, it will, whenever it brings about an effect, bring about E. But I do not accept it if it means that whenever C occurs, it is necessary that E occur.

[17] I have spelled out this claim in "Intentionalists and Davidson on Rational Explanations" in G. Meggle, ed., *Actions, Norms and Values: Discussions with Georg Henrik von Wright* (Walter de Gruyter, 1999), pp. 191-208.

[18] Cf. my "Responsive Action and the Belief-Desire Model"

[19] See his "Thinking Causes" in *Truth, Language and History* (Oxford University Press).

IV

Sehon's main defense of his claim that CSP explanations are not causal is that if they were, CSP would be reducible to physical science, a defense he makes a number of times. For example, he defends the claim that "If beliefs can causally explain behavior, then beliefs are brain states" by arguing for its contrapositive: "If beliefs are not brain states, then they cannot causally explain behavior." He then argues that beliefs can neither be reduced to nor be token identical with brain states, which, given the contrapositive, entails that beliefs cannot *causally* explain behavior. But since beliefs *do* explain behavior, they must do so non-causally.[75] Later he argues against the view that an agent's behavior might have two different causes – a physical cause and a mental cause such as a desire. His main objection is that "if human physiology ultimately gives a gapless causal history of bodily motions, either mental states are identical to the physiological cause or they are not causes of behavior,"[202] but since mental states are not identical to physical states, they are not causes of behavior.

This defense of the non-causal nature of CSP explanations assumes that causal explanations are found only in physical science – that only physical explanations are productive. Why that assumption? Some make it because they think that only in physical science do we get the true knowledge required for genuine explanation. Steven Pinker, for example, writes that CSP "has so much power and precision in predicting, controlling and explaining everyday behavior… that the odds are high that it will be incorporated in some form into our best scientific theories." [Quoted by Sehon, 216] But Sehon does not accept this reductionism because "there are truths of common sense psychology that are logically independent of, and hence not explained, by the truths of physical science."

He does, however, accept (without defending) the causal completeness of physics, which he takes to be decisive for his view. He accepts, that is, the claim that all physical effects are due to physical causes or, as David Papineau puts it, "All physical effects are fully determined by law by a purely physical prior history."[20] While many who accept this (empirical) claim think it entails reductionism of some kind, Sehon does not. He maintains that there are truths of CSP (along with terms, properties, and entities) that are not reducible to physical science. While he does contend that "we are constituted by elementary physical particles [as is] anything that is capable of having effects," which is everything except "numbers and many other abstract objects," [133] he does not count that as reduction. What he thinks does follow from the causal completeness of physics is that whatever causally explains our behavior must be physical. Physical science "will ultimately provide a gapless causal history [of behavior], a history that appeals only to physical states of the agent" [201], which he takes to entail that there are no causal explanations in CSP.

But the latter does not follow from the causal completeness of physics. What does follow is that causal explanations in CSP are not a part of physical science – that such causes do not function to fill in gaps in the causal histories the latter constructs. Papineau argues that what made the causal completeness of physics so plausible in our time is that physics was able to establish that "there is one quantity, energy, preserved in all natural interactions whatever," which rules out any non-physical forces (vital, mental, etc.) that do not reduce to "fundamental conservative forces" and hence enables physics to "uphold the universal conservation of energy." If this argument is decisive, it follows that explanations in CSP do not appeal either to the fundamental forces of physics or to any special forces

[20] "Appendix" to his *Thinking About Consciousness* (Oxford, 2002), p. 250.

over and above them.[21] But it does not follow that such explanations are not causal; causal explanations in CSP are not a matter of special forces, whether or not they are reducible to "a small stock of fundamental forces." Even if physics is causally complete, there may be causal explanations of a quite different type.

My view is, therefore, that two different types of causal explanation of human behavior are acceptable, one belonging to and serving the purposes of physical science, the other belonging to and serving the purposes of CSP. Two main objections will be raised to this view. One is that it is not consistent with the completeness of physics, the other that there cannot be two causal explanations of the same behavior.

The first contends that if one accepts the claim that all physical effects are due to physical causes, then a CSP causal explanation outside physical science entails either that some physical effects are due to non-physical causes or that CSP explains non-physical effects. In either case, one appears committed to the kind of dualism ruled out by the completeness of physics – a dualism between the observable physical and the introspectable mental.

This argument, however, equivocates on "physical". The claim that physics is complete uses the term in a *physicalistic* sense: it does not mean by "physical effects" the middle-size observable phenomena that are part of our everyday world (including human behavior) but the carefully described phenomena that figure in the experiments designed to test the theories of physics. The same holds for "physical causes", which are not the observable states of affairs that agents cite as reasons in giving rational explanations of why they act but the theoretically characterized physicalistic entities and processes of physics. The explanations of CSP, therefore, appeal to non-physical causes only if that means non-*physicalistic* causes, and they explain non-physical effects only if that means non-*physicalistic* effects. But those explanations do not appeal to causes and effects that are non-physical in the sense of being immaterial, not in space, or not knowable on the basis of observation. We do not, therefore, have to choose between physicalism and dualism.

This distinction between the technical physicalistic and the everyday physical does not mean they exist in different worlds since it is a *conceptual* distinction between two ways of describing (conceptualizing) phenomena – either as in physics or as in CSP. The two ways of describing phenomena yield two corresponding types of explanation since explanation is always of phenomena *as* described. Explanation, that is to say, is intensional: from the fact that E explains x and x=y, it does not follow that E explains y. This is generally recognized in the case of intentional action: that the beggar needed food explains Klaus's action described as "intentionally giving her money" but not as "causing a small riot in the street," even though both descriptions are true of the same act. It is often overlooked that this point also applies to explanation generally. Something falls from a high window: physics explains the phenomenon *as* a falling object by appealing to the law of gravity, but it does not explain that same event *as* someone jumping out of a window, for the latter description is not part of any natural law.

This distinction is especially pertinent to the explanations of behavior that Sehon discusses. He makes the point that "notions like *purpose*, *goal direction, belief*, and *desire* have no role in physical science" but he does not make the corresponding point about the behavior these notions are supposed to explain, namely, that *as described in terms of CSP*, human behavior has no role in physical science. "Human behavior" is like "physical" in

[21] I do not accept the causal completeness of physics and hence do not agree that arguments like these establish it. The main problem is that these arguments mistakenly assume that the *interaction* of the fundamental forces of physics conform to the same laws as the forces taken separately.

having a technical sense in physical science *and* an everyday sense in CSP. A neuroscientific account of human behavior does not explain behavior *as* described in CSP, *as* walking to the kitchen or waving to a friend; it explains it described *as* "colorless movements," which are specialized and abstractive descriptions that leave out distinctive CSP features of behavior. Sehon's assertion that a "brain scanner will be [relevant] only when our concern is directly related to the person's motor behavior and its physical causes" is correct if "motor behavior" is used as neuroscientists use it and "physical causes" means "physicalistic causes". But CSP is also concerned with motor behavior *when* it is described *as* an agent's moving her body and limbs and thereby intentionally doing various things, for that puts motor behavior in the domain of CSP, not as colorless movements but as an agent acting.

The second objection to my claim about two types of causal explanation of human behavior concerns cases where the same (motor) behavior is explained in CSP (someone moves her body and limbs in order to pull a rope) and in neuroscience (her bodily behavior described as "colorless movements"). Jaegwon Kim puts the objection as follows: "A 'purposive' explanation of human action in terms of the agent's 'reasons' and a 'mechanistic' (e.g. neurobiological) explanation of it in terms of physiological mechanism must be regarded as incompatible and mutually exclusionary – *unless* we accept an appropriate reductive relationship between intentional states and underlying biological processes." His objection appeals to what he calls the "principle of explanatory exclusion": "there can be no more than one 'complete' and 'independent' explanation for any single explanandum."[22]

I believe Kim is mistaken. Let us assume that the two types of explanation are *independent*, and hence that explanations in CSP cannot be reduced to "underlying biological processes." Let us also assume that each is *complete* in that one makes it intelligible why the agent acted intentionally as he did, the other why his physicalistically described bodily movements occurred as they did. My view does not violate Kim's "principle of explanatory exclusion" since if we have two such independent and complete explanations of the same behavior, the explanations will be of different types. This implies that, although we explain the same behavior, we do not have, as far as Kim's principle is concerned, a single *explanandum* because (as argued above) in giving an explanation of behavior, we must explain it *as* described, which in this case means either *as* described in CSP or *as* described in neuroscience. These different types of description of an agent's behavior are compatible, and since explanation is of behavior *as* described, the explanations are also compatible.[23]

V

My contention that causal explanations of a distinctive type are central to CSP is often admitted by philosophers who, granting that CSP explanations are distinctive, nevertheless think they cannot stand by themselves but must be supported by external explanations from

[22] Jaegwon Kim, *Supervenience and Mind* (Cambridge University Press, 1993), p. xiii.

[23] The notion of "same behavior" calls for more reflection since the two types of description do not, in general, *individuate* in the same way. But if what is differently described is not in the strict sense the *same* behavior, then Kim's principle is preserved at another level. The best thing to say, however, is that there is no behavior over and above what is described in one way or another: "the *same* behavior" is not a determinate description of behavior but a determinable description made determinate only by descriptions from CSP, neuroscience, etc.

physical science. While Sehon does not think CSP explanations are causal, he also thinks they cannot stand by themselves.

> …If teleology is irreducible, then we have no explanation for why teleological explanation works…. If we had a causal analysis of teleology…. we would expect that cognitive science would find the causal story underlying the cognition of humans and other animals, and that this causal story would explain the applicability and legitimacy of teleological concepts.[172]

Sehon simply takes it for granted that CSP needs support from physical science, as do most philosophers who agree with the claim. The most explicit attempt to defend it that I know of is Hartry Field's paper on "Physicalism."[24] Field begins his paper by asserting that "I take it as beyond serious doubt that there is an important sense in which all facts depend on physical facts and all good causal explanations depend on good physical explanations." That is a rough statement of what he means by "physicalism," which he does not give a great deal of argument since he regards it as beyond serious doubt. His main defense is that "some such doctrine has played an important methodological role in guiding the development of science." He spells that out as follows:

> The methodological role of the doctrine of physicalism is double-edged. On the positive side, the doctrine tells us that when we have a putative body of facts and causal explanations that we are quite convinced are basically correct, we need to find a physical foundation for them…. For instance, the implicit acceptance of the doctrine of physicalism on the part of most scientists has led to the successful search for the molecular foundations of genetics and the quantum-mechanical foundations of chemical bonding. The other, negative, aspect of the doctrine of physicalism is that when faced with a body of doctrine (or a body of purported causal explanations) that we are convinced can have no physical foundation, we tend to reject that body of doctrine (or of purported causal explanations). I think this is the attitude that most of us take toward astrology or telepathy: even if there were positive evidence for telepathy that we did not know how to refute, most of us would tend to disbelieve the telepathic claims (and presumably suspect the evidence) simply because it seems so difficult to conceive how such claims could fit in with a physicalistic worldview.[25]

There are two main objections to this argument. The first is that, whether or not physicalism has played an important methodological role in guiding the development of physical science,[26] CSP is not a physical science, and it is absurd to think of its developing by finding a physical foundation for its claims. Theories about dreaming, mental illness, personality disorders, and other psychological phenomena have developed over time, but it is no part of CSP to articulate such theories. Field thinks it is because he thinks CSP is a

[24] In *Inference, Explanation, and Other Frustrations*, ed. by John Earman (University of California Press, 1992)

[25] "Physicalism" 271f.

[26] I do not think that the doctrine of physicalism has played the guiding role in the physical sciences that Field assigns to it. Chemists and biologists seek explanations from physics for various chemical and biological phenomena, but many of them reject the reductionism that Field takes to be integral to physicalism. They reject it because they think a robust sense of the reality of distinctively chemical and biological phenomena is essential for guiding the quest for underlying physical explanations. The quantum mechanical explanation of chemical bonding, for example, presupposes the notion of a chemical kind, which is a macroscopic phenomenon not reducible to physics. See, for example, Paul Needham's discussion of reductionism in his *Law and Order* (Stockholm Studies in Philosophy, 2005)

"special science" whose "first explanatory task is simply to explain in terms of an underlying science like physics why generalizations of this theory should hold."[283] But CSP is not a special science, and since its rational explanations of intentional action are not based on generalizations, it is irrelevant to seek a physical explanation of why they hold. The essential task of CSP is to describe and explain intentional action, which is something that does not depend on knowledge that has developed over time in any relevant sense.

A second objection is to what Field calls the negative aspect of physicalism: to accept causal explanations that have no physical foundation would be like accepting the claims of astrology or telepathy. We disbelieve the latter, he holds, "simply because it seems so difficult to conceive how such claims could fit in with a physicalistic worldview,"[27] so that rejecting physicalism would open the door to all sorts of pseudo-scientific claims and theories.

But one surely does not have to accept physicalism to maintain that astrology and telepathy are contrary to the knowledge provided by physical science. It might be true, as Field suggests, that we are unable to *refute* such pseudo-sciences unless we appeal to physicalism, but we may also be unable to *refute* clever flat-earth believers, and they can be ardent physicalists. Moreover, the strongest objections to such pseudo-sciences are provided by special sciences, and they are irreducible to physics – irreducible, that is, unless one accepts strong physicalism, which would beg the question at issue. In any case, CSP, unlike astrology or telepathy, does not purport to be a science. Sane human beings may, of course, give astrological or telepathic reasons for their behavior. But to reject them as normatively justified reasons is not to deny that an agent's acceptance of them may be integral to a causal explanation of why he acted as he did.

Field's skepticism about the validity of the latter type of causal explanations – that "we tend to reject…purported causal explanations [that] have no physical foundation" – assumes that, absent a physical foundation, we are unable to distinguish acceptable from unacceptable rational explanations. Establishing a causal explanation in CSP is, of course, different from establishing one in physical science: appeal to generalizations or experimental manipulation is irrelevant, and while appeal to scientific data is often relevant to whether a reason is normatively acceptable, it is not usually relevant to whether a reason is explanatorily acceptable. Indeed, the CSP explanations an agent gives of her own behavior are not based on observation or evidence of any kind. Such explanations are not, however, incorrigible, and they may be challenged by other agents, who do appeal to evidence of various kinds. The latter may be based on observation of what the agent did before or after the action in question, on her character and what she could or could not have done, on the situation in which she acted, on her past life, and so on. Considerations of that sort are, in general, quite sufficient to distinguish acceptable from unacceptable rational explanations of an agent's behavior.

These are *interpretative* considerations of the kind that are essential to the ascription to an agent of actions, mental states, and reasons, which means their ascription is holistic. While explanations in physical science may be holistic in the Duhemian sense of involving complex trade-offs between theory and observation in the explanatory process itself, they are not holistic in having to take account of the complex trade-offs in *what* is being explained, where an explanation of the agent's action is validated by showing that it makes maximum coherent sense of her actions, mental states, and reasons for acting. That open-ended process is necessary for confirming a causal explanation in CSP, and it may involve indeterminacy in our conclusions. There is nothing like this in physical science where,

[27] "Physicalism" 271f.

however tentative a conclusion may be, the aim is to articulate it determinately by giving a precise statement of how things are.[28] CSP explanations may occasionally leave it open whether an agent acted for one reason rather than another, not because we do not know enough to decide but because there is no answer to the question. It does not follow that rational explanations are defective, only that they are different.

Field offers a further argument for the necessity of a physical foundation for CSP, namely, "not to explain the laws of the special sciences themselves but simply to explain why the application of the special-science laws never comes into conflict with the application of the underlying laws."

> This is in effect a demand that we explain why our neurophysiological laws and our psychological laws never come into conflict. Or, to introduce a convenient phrase, it is a demand that we show that our neurophysiology and our psychology "mesh." It seems to me that whenever we employ laws at different levels, there is a prima facie possibility of their coming into conflict, and it is eminently reasonable to want an explanation of why such conflict does not arise. I take it that a main advantage of reducing psychology to lower-level science… is that doing so one would be able to explain the mesh between psychology and the lower-level sciences. [285]

In putting this point, Field assumes that CSP is a special science that aims to establish general laws, an assumption I have rejected. But let us waive that and modify his point so that it asks for an explanation of why neurophysiological and CSP explanations never come into conflict – why they "mesh."[29]

I have discussed this point at length elsewhere[30] and hence will only summarize my view here. Given the modified point, what Field calls "mesh" (which I call "congruence") concerns the relation between a rational explanation of an agent's intentional behavior and a neurophysiological explanation of the bodily movements that are involved in that behavior. Field thinks we need a substantive explanation of that relation and suggests reduction of CSP to a lower-level science. On my view, a substantive explanation is simply out of place. What is supposed to be explained is why the movements of, say, Mary's right arm in her intentionally moving it (for example, to pull on a string) mesh with the movements of her arm as described in neurophysiology. But there is neither need nor place for giving a substantive explanation of *that* because those are the same movements, although differently described – in terms of CSP, on the one hand, in terms of neurophysiology, on the other.

If Mary moved her right arm to pull on the string, then a neurophysiological explanation of her arm movements *could not fail* to be an explanation of the movements involved in her having intentionally moved her right arm since the arm movements explained (in CSP) by her intentionally *moving* her arm *are* the movements we explain neuroscientifically. The movements are described differently, but what neurophysiology describes in a specialized and abstractive way *are* the movements Mary made in moving her arm in order to pull on the string.

[28] Even when the conclusion is probabilistic in form, the aim is a precise quantitative statement of the probabilities, not a statement whose indeterminacy means there is no fact of the matter.
[29] Sehon raises a similar problem on p.216, which he seems to resolve by his claim that teleological and causal explanations are logically independent and answer very different questions. This, however, overlooks the point that CSP does not simply describe behavior in teleological terms but also explains what produces it.
[30] "The Problem of Congruence" in Niiniluoto and Vilko, eds., *Philosophical Essays in Memorian: Georg Henrik von Wright (Acta Philosophica Fennica*, Vol. 77, 2005).

The movements of an agent's body and (what we take to be) her intentional action *can* fall apart. If Mary had set out to turn on the lamp but her upper body had (unknowingly) become paralyzed, she would have been unable to move her body so as to pull on the string. But the issue of mesh would then be moot, because her behavior would not be her intentionally pulling on the string, or even her intentionally moving her arm. It might have presented itself as having the *form* of intentional action but it could be explained *only* as bodily movement in neurophysiology.

Consider Mary as unable for some reason to move her right arm and hence unable to pull the string as she usually does. She would then use her left arm, and a CSP explanation of her acting would refer to her moving that arm to pull the string. The movements she thus made could also be described in terms drawn from neurophysiology and (let us assume) be given a neurophysiological explanation. But what the latter explained would clearly have to be the same movements Mary intentionally brought about in moving her left arm. The explanations drawn from CSP and from neurophysiology would mesh, therefore, because the movements each explained would be the same movements differently described.

In explaining an agent's behavior, therefore, there is *necessarily* a mesh between the two types of explanation. If Mary failed to act intentionally as expected because she could not move her limbs in the way required for *that* action but was able to act in *another* way by making different movements, then the movements that resulted from her intentionally moving her body and limbs would also be different. But so would the neurophysiological descriptions of her behavior since they describe those same movements (in neurophysiological terms), and their neurophysiological explanation would be adequate only if it yielded the movements as thus described.

Field would likely reject this account of meshing because it takes CSP explanations as basic and requires that neurophysiological explanations conform to them. That gives CSP explanations priority over neurophysiological ones, which is unacceptable, it will be objected, to anyone sympathetic to the achievement and status of physical sciences like neurophysiology.

My response is that neither type of explanation is basic in an *overall* sense because which type has priority depends on the context and the questions being asked. When the question concerns their mesh CSP explanation takes priority. The reason is that to consider the mesh between the two types of explanation of the movements of the body involved in acting, we must identify which movements they are. When an agent acts, his body moves in all sorts of ways, many of which are not relevant to what he is doing intentionally. To identify the movements that are relevant, we must identify his intentional act, and that requires putting his acting in the context of CSP and, typically (for holistic reasons), identifying the reason for which he acts. It is *those* movements, the ones identified by their role in a CSP explanation, that we also aim to explain (though described differently) neurophysiologically.

There are other contexts in which neurophysiological explanations are prior, for instance, explaining why agents are *incapable* of certain actions. CSP explanations cannot explain such incapacity because they presuppose that the agent is *capable* of acting and hence that he is normal. To the extent that he is normal, however, rational explanation takes priority, which means that there will *necessarily* be a mesh between his behavior as action and as mere movements. We can, therefore, say that while abnormality can be substantively explained, normality cannot. It does not follow that we cannot explain why beings with the potential to become normal agents have come to exist – because of

evolutionary considerations or more short term explanations that might belong either to physical science or to CSP.

Of course, such beings might never have come to exist, and then there would have been no behavior to be explained in CSP. Such an impoverished world illustrates another context in which neuroscientific explanations are basic, namely, what would have been explained had there been explainers can exist without normal agents, but not vice versa. That is one way of expressing *global* supervenience of CSP on physical science: destroy the physicalistic and you destroy everything else, while the contrary is not true. I accept that, but it has no consequences for the nature and function of CSP.

Davidson's view of whether CSP explanations need a physical foundation is subtle and merits further discussion. It is widely thought that his "Principle of the Nomological Character of Causality" is his way of providing such a foundation. That principle means, not that causal explanations are in terms of general laws (which Davidson denies), but that any causal explanation entails that there is a causal *relation* between events related to the explanation, which in turn entails that those events have descriptions that are instances of a strict law, hence a law of physics. This conception of the relation of causes and laws was thought to be Davidson's way of grounding CSP explanations in physical laws because he was construed as holding that causal explanations in CSP had their causal force *in virtue of* the laws of physics they entailed. But Davidson denied ever claiming that: on his view, a causal relation between two events *entails* that there is a physical law connecting physical descriptions of the two events but this does not mean they are causally related *in virtue of* such a law.

Davidson gave no arguments for his "Principle of the Nomological Character of Causality" until his 1995 paper on "Laws and Causes." He argued there that causal explanations explain only changes (events) and that what is a change is relative to how a situation is described (which he illustrates with Goodman's points about green, grue, blue, and bleen). Since descriptions of what is a change must involve law-like predicates, it follows that there are causal explanation only when there are laws. That is sufficient, he maintains, to show that "singular causal statements… entail the existence of strict laws."[31]

That description does not do justice to his subtle paper but it is enough to show that he asserted the cause-law connection because of conceptual relations between laws, changes, and causes, not because the connection provides a physical foundation for CSP explanations. This reinforces the point that Davidson's view is that causal explanations *entail* the existence of physical laws, not that they are valid in virtue of them, which means that Davidson, unlike Field, does not think that "all good causal explanations depend on good physical explanations." The fact, moreover, that we need not know what physical laws causal explanations entail also bolsters the point that they do not support CSP, for laws of which we are ignorant cannot increase the power of a causal explanation to render intelligible why an agent acts as she does.

VI

In this final section I will defend the claim that CSP explanations can stand alone without external support. I agree with Sehon that it is "a brute, irreducible fact with no further explanation" that the principles of CSP "hold reliably of human beings and other agents [219] *if* that means that CSP requires no *external* support, but not if it means that the success of CSP is unintelligible, inexplicable, or a mystery. Reflection on the nature of CSP

[31] Now in *Truth, Language and History*, 219.

itself can show why it works, and hence we need not leave unanswered the question of why its explanatory claims are true and its explanations successful.

We should first consider the background to this question – the assumptions made in posing it. Sehon quotes [216] Steven Pinker as writing that: "…Scientific psychology will have to explain how a hunk of matter, such as a human being, can have beliefs and desires and how the beliefs and desires work so well." This is a very misleading way of posing the question about why or how CSP works, and, although Sehon rejects Pinker's answer, he accepts his way of posing the question.

> We are constituted by material particles, and these material particles don't suddenly cease to follow the laws of nature just because they are embedded in the body of an agent…. How can it even be possible that there are nonphysical facts about physical objects? Given that we are physical objects, how is it even consistent to maintain that there are facts about us that do not reduce to physical facts? [231]

To ask, "How can it even be possible that there are nonphysical facts about physical objects?" is to ask "How is it possible that there are CSP facts about *physicalistic* objects" since the latter are what hunks of matter or physical particles are. However, while there are true descriptions of human beings as hunks of matter or physical particles, it is misleading to say that is what human beings *are* (or *are constituted by*), since it misses completely what is distinctive about them. Moreover, many facts of CSP are physical in an everyday sense since they are about human behavior, which is quite physical; they are nonphysical only in the sense of being non-*physicalistic*.

Physicalistic descriptions of human beings characterize them in an idealized sense since they ignore the concrete contexts in which they live and act, and they describe them abstractively in that the physical forces involved are characterized in abstraction from the innumerable ways in which the latter interact.[32] To ask how there can be facts from CSP that are about hunks of matter or physical particles is, therefore, to begin with idealized and abstractive descriptions of human beings and then ask how described in *that* way, they can also be described as CSP does. In one sense, the answer is obvious since what CSP describes in its own concrete (normative and agent-centered) ways just *is* what physical science describes in its idealized and abstractive ways. But Pinker or Sehon do not want an *obvious* answer; they want to know how what is described *only* as hunks of matter or physical particles can also be described as CSP does and successfully explained in its terms. There is no answer to *that* because it is not a good question. The agents whose behavior (thoughts, feelings) we explain in CSP do not *consist* of hunks of matter or physicalistic particles, since that is at best a *physicalistic* way of describing them. Of course, such agents *are* physical but in the everyday sense that is not identical with the physicalistic. They also exist in a world that is physical in that everyday sense, and they have many causal powers that are not adequately characterized in physicalistic terms. It is facts about agents in *that* sense "that do not reduce to *physicalistic* facts," and it is the success of CSP explanations of the behavior of agents in *that* sense that we are trying to understand.

We can ask why there are such agents at all, a question that has an initial Darwinian answer: there are clear survival benefits to beings that evolve so as to be increasingly

[32] Nancy Cartwright writes that "the rules of composition [of laws of nature] are empirically supported… only so long as nothing *interferes*…. Our first order principles and our principles of composition support only claims about what happens so long as all relevant factors can be correctly described with the theory….' [Discussion of *The Dappled World* in *Philosophical Books* (October, 2002), p. 243.] "Interference" is, of course, central to CSP, which has many relevant features that cannot be correctly described within physicalistic theory.

capable of behavior that can be explained in CSP. We can then explain how individual beings, who are born of parents that have thus evolved, themselves become mature human agents: they have, relative to other animals, a long maturation period, which permits their being trained and educated in and by the human community to give and receive reasons for their action. This capacity to act for reasons is sustained and further developed by the fact that reasons for acting are embedded in various practices, institutions, and artifacts and by the fact that participation in the latter affects the structure of the brain and nervous system.

What is thus explained is the existence of beings with the capacities of intentional agents. While that does not explain why CSP explanations are successful, it shows that what is to be explained is not how hunks of matter or physical particles are able to give successful CSP explanations of each other but how *human agents* can do so, which is the way the question should be posed.

Pinker speaks of CSP as having "so much power and precision in predicting, controlling, and explaining everyday behavior," which he illustrates by describing two persons who agree to meet at a bar in Chicago at a certain time two months hence and do just that: "That is amazing! In what other domain could laypeople – or scientists for that matter – predict, months in advance, the trajectories of two objects thousands of miles apart to an accuracy of inches and minutes.?" [quoted on 216] This is, however, the wrong way to characterize the success of CSP predictions because it uses the terms of physical science. CSP predicts, not the trajectories of two objects thousands of miles apart, but the intentional actions of two agents who have communicated with each other. It predicts what each will do intentionally and the reason why, but it makes only very vague predictions about the movements of their bodies and limbs, which can vary widely as long as they are sufficient for their actions. The "power and precision" that CSP has must be characterized in the language of CSP itself, and while it falls far short of physical science in predicting the trajectories of bodies, it far exceeds it in predicting what agents will do intentionally (or at least try to do).

At the same time, there are many actions that CSP cannot predict if for no other reason than that agents often change their minds. Moreover, even if an agent does not change his mind and we can predict that he will be at a bar at a certain time and place, not only can we not predict the movements of his body and limbs except very vaguely, but we cannot predict how he will get there or what his going will lead to. Indeed, predicting what an agent will accomplish is often less important than explaining what he is trying to do and why. Pinker speaks of our controlling everyday behavior, which we can, of course, often do, but we do so on a very different basis than we control the trajectory of a missile, the course of a river, or color of a substance. We cannot control an agent's *intentional* behavior by forcing his action or deceiving him, for what we thus control he does not do intentionally. To control what he does intentionally requires that we link up with the reasons that explain why he acts, and that requires that we explain his behavior in terms of CSP.

The deepest measure of success for CSP explanation is that it enables us to deal with each other as human beings. It enables us to evaluate agents and their actions and to hold them responsible for what they do, crediting them, if we wish, for what they do well, blaming them, if we must, for what they do badly. It enables us to recognize the range of emotions people express in their actions, to know when they are suffering or when they are pleased, to respond to what they are feeling and intending and not merely to their external movements. It enables us to work together, to respond intelligently to needs and desires, to

cooperate in making the space and objects in which we live. The success of CSP, in short, is not specifiable apart from accounts of what it is to live a human life.

The first reason I suggest for why CSP explanation is successful in this way is that it is constitutive: to be a human agent capable of acting intentionally *is* to be capable of giving and receiving successful CSP explanations of the actions of oneself and others. As an account of the success of CSP this is, admittedly, very thin because what is to be accounted for – the success of CSP explanations – is identical with what accounts for it – namely, an agent's capacity to act intentionally. It is a kind of elucidation, however, since its point is that to give an account of why CSP explanations are successful is, in part, just to give an account of what it is to live a human life. CSP explanations are not successful because they meet an external standard; they are successful because there would otherwise be no such thing as human existence. Some may argue that human existence itself has an external purpose; even if that is the case, it is not the standard by which to judge the success of CSP explanations. They enable us to live as human beings, not to fulfill some external end of human existence.

To alter the character of CSP explanations in a fundamental way would be, therefore, to alter the character of human life itself. Proposals to reform CSP that are based on external standards would also be proposals to change the nature of human life, changes that are based on external standards. Such a deep reform of human life is not impossible but it would have to come from within human life itself and hence from within CSP. Proposals from physical science for such a fundamental reform could conceivably be accepted, but only to the extent that they managed to effect a change not only in CSP but in human existence itself.

The second reason I suggest is that CSP explanation is successful in that it works *among us*. The capacities we have acquired as agents through evolution, culture, and education are capacities to coordinate our lives so that we can act, feel, and think together. Crucial to this are the commitments we give and acquire in making intentions or giving promises, in fulfilling roles (as parents, teachers, workers, friends, etc.), in making contracts, in borrowing and lending, and so on. Such commitments enable us to predict what others will do, to plan what we will do, to settle matters so that we do not constantly have to make new decisions or deal with the always new decisions of others. CSP works so well because we constantly engage, explicitly or implicitly, in making these commitments. (It explains, of course, why Pinker's two men were able to meet in a bar two months after a conversation.) If it is asked why we make them, various answers may be proposed: we were brought up that way, that is the way life goes on in our society, that is the very fabric of human existence. If it is asked why we keep such commitments, similar answers are relevant, along with the fact that to understand what it is to make such commitments is to understand that one will fulfill them – except under certain conditions that are also mutually understood.

One may push deeper, however, and ask what underlies the success of these commitments in enabling us to act, feel, and think together, especially given that it was such success that enabled us to acquire the capacity to give and receive them in the first place. The basic answer is that we live in a common world that provides common reasons for our action that we can perceive in common. This obvious truth rules out the notion that we are in touch with the world only through getting information about it – that what we encounter directly is a virtual reality that is in each of our heads and needs to be coordinated. If that is ruled out, then our encounter with the common world (however that is explained) is

sufficient to explain our ability to make and keep the mutual commitments that support our common acting, feeling, and thinking.

This shows that CSP's primary virtue is not truth but its enabling us to live and cope with each other and with the world in which we are embedded. It is, if you will, a way of being in the world. But that does not mean there are no truths in CSP. I agree with Sehon that "There are truths of CSP that are logically independent of, and hence not explained by, the truths of physical science," but I disagree with his going on to write that "accordingly, these truths will apparently not be susceptible of further explanation; thus in addition to whatever mysteries physical science leaves, CSP will introduce further inexplicable mysteries." [215] "By virtue of what are [the principles of CSP] true of us?" he asks, answering that "Part of what it is to have a nonreductionist theory of mind is to have questions like these left unanswered." [231]

The notion of "that in virtue of which a proposition (or principle) is true" in the sense of what *explains* why it is true – *makes* it true – seems to me confused. The question "in virtue of what is a proposition true?" can be answered only in Tarskian fashion: "'p' is true if and only if p," which is not an *explanation* of why 'p' is true (or what makes it true) but an account of what it is to be true. The reason we cannot explain why the principles of CSP are true is not that they are irreducible; the Tarski point applies to any proposition, not only to those that belong to a nonreductionist theory.[33] There would be, in any case, no explanatory force in saying they were made true by physicalist facts since that would just be a way of saying that they are physicalist truths (which Sehon denies).

This does not mean that the truths of CSP are to be construed in an anti-realist, perhaps instrumentalist, way. There are intentional actions, reasons for action, beliefs, desires, and intentions, all of which are just as real as any physicalistic entities. The same is true of the everyday physical world, whose macro-entities do not have second class reality compared to the particles of micro-physics.[34]

We can, of course, specify what makes the claims of CSP true if that means what *evidence* there is for them. Claims of CSP are false if they specify the wrong reason for an agent's action, misdescribe what she has done, or ascribe to her a belief she does not hold or a desire she does not have. But there is typically, in principle, sufficient evidence to determine when such claims are false and to correct them because, as Sehon puts it, "CSP is constrained by its own internal principle…." [231] We can distinguish between the reason for which an agent acted and what merely appears to be her reason, between what she really did and simply claims to have done, between what she did and did not believe. These are interpretive and not scientific claims, but that is what we want and what we get from the explanatory truths of CSP.

[33] This point is defended by Frege and Davidson, and I defend it in my "What Philosophers Should Know about Truth and the Slingshot" in Sintonen, Ylikoski, and Miller, eds., *Realism in Action* (Kluver, 2003), pp. 3-32.

[34] I agree with Cartwright's claim that "concepts from macrophysics and from various branches of technology and engineering are required *in conjunction with* those of 'microphysics' to obtain true law statements…." [Cf note 32 above]