Interactionism for the Discerning Mind?

Derek Shiller

This is a post-peer-review, pre-copyedit version of an article published Philosophical Studies. The final authenticated version is available online at:

https://doi.org/10.1007/s11098-018-1213-5

Doubts about the possibility of causal interaction between non-physical minds and physical brains go back to the birth of interactionist dualism. Jaegwon Kim (2001; 2005) has recently tried to clarify the concern.[1] According to Kim, interactionists face a challenge in accounting for the distinct causal capacities of different minds. Common sense holds that particular minds control particular brains: my mind can cause my body to go into the kitchen and fetch a cup of tea, but it cannot do the same to yours. Interactionists must make room for these differences in causal powers by explaining how is it that minds might be causally paired with only the bodies they influence. Kim argues that interactionists cannot succeed. This has come to be known as the 'pairing problem'.

Interactionists have offered a variety of responses to the pairing problem. In this paper, I will argue that the problem is deeper than its critics have acknowledged. While Kim focused on the capacity of minds to influence specific brains, exercising this capacity requires influencing specific neurons. The neural discernment problem is the pairing problem writ small: the difficulties interactionists face in explaining mind-brain pairing are mirrored by difficulties in explaining mind-neuron pairing. I will argue that the neural discernment problem adds complications that frustrate the best responses to the original problem.

In the first section, I lay out my favored explication of the pairing problem. In the second section, I will present the neural discernment problem. In the third section, I discuss four responses to the pairing problem – pairing with spatial relations, pairing with union relations, pairing with intrinsic properties, and brute mind-brain causation – and explain how each response fairs worse when set against the neural discernment problem.

[1] Kim attributes his worry originally to Foster (1968).

The Pairing Problem

To be plausible, interactionist theories need to make sense of our minds' capacities to influence and in turn be influenced by specific brains. Given the current epistemological limitations of neuroscience, interactionists cannot yet be expected to work out a theory of mind-brain interaction. That said, we do know something about how the brain works, and so interactionists need to make it plausible that mind-body interaction is consistent with some elaboration of our present understanding. If interactionism makes the causal pairing between minds and brains miraculous, the view is untenable.

The original pairing problem depends upon two premises.

The first premise is that differences in causal powers require non-causal differences: it is problematic for a theory to posit causal differences that cannot be explained by means of available non-causal differences.

There are multiple ways of clarify this idea;[2] the formalization I favor, the *causal accountability principle*, says that differences in the causal power relations of pairs of objects (i.e. x has the power to produce change y in z) always depend upon appropriate non-causal differences.

> Causal Accountability Principle: If two objects differ in their causal powers, they do so at least in part by virtue of differing in their intrinsic properties or in the relations that they bear to the subjects of those powers.

Kim adopts a metaphysical interpretation of this idea, according to which it is metaphysically impossible for two objects to differ in only their causal powers. This interpretation allows him to draw a strong conclusion: if interactionists cannot locate differences in the intrinsic properties of minds or the relations

[2] Kim spends little space elaborating on this requirement, and so he might not accept my formulation of the relevant principle. It differs slightly from what other commentators have provided. Audi (2011) uses "if two things interact causally, they instantiate a pairing relation" (4). Bailey, Rasmussen, and Horn (2011) use "for all x and all y, if x causes y, then there is a relation or relations... such that their holding makes it the case that x causes y" (350). Saad uses "necessarily, objects differ causally only if they differ non-causally" (Saad 2017) and "necessarily, causal differences between objects are at least partly grounded in non-causal differences between them" (Saad forthcoming) .

One notable difference between my formulation and these others is that mine concerns causal powers rather than causal relations. A causal power encapsulates an object's (probabilistic) potential to produce causal effects in specific situations. An object has the causal power to produce an effect in a given situation if it makes it more likely to occur in that situation. Causal powers are intended as overt behavioral properties and not inherent dispositional properties (considered as standing properties that are characterized by their capacity to explain dispositional behavior), for causal powers may be explained by such properties. The advantage of formulating the principle in terms of causal powers is that it allows for unaccountable differences in the chancy instantiation of causal relations. (Saad forthcoming) develops a related proposal.

that they bear to their brains, then they cannot attribute them different causal powers.

The second premise is that neither minds' intrinsic properties nor their relations are able to account for the differences in their causal powers. In order to accommodate the possibility of physically indistinguishable but distinct brains that are separately controlled by psychologically identical but distinct minds (as would be the case if we had perfect duplicates somewhere else in the universe), pairing cannot be achieved by intrinsic properties.[3]

Kim suggests that this leaves only a few plausible candidate pairing relations. In order to take advantage of the spatially distinctive differences between brains (the only differences that distinguish physically identical brains) the relevant mind-brain relations require a space-like structure. Here, however, interactionists are faced with a dilemma. Either they must posit novel relations specifically for the purpose of pairing, or they must locate minds in space.

Kim denies that interactionists can locate minds in space on the grounds that it will either result in absurdities or else will collapse the view into physicalism.[4] Furthermore, he regards the maneuver of positing new relations specifically to make sense of pairing as unacceptably ad hoc and explanitorily shallow.

We can formalize Kim's argument as follows:

1. If two objects differ in their causal powers, they do so at least in part by virtue of differing in their intrinsic properties or in the relations that they bear to the subjects of those powers.
2. If interactionism is true, then intrinsically identical minds can differ in their power to influence intrinsically identical brains.
3. If interactionism is true, then minds bear no relationships to their brains that are appropriate[5] for accounting for causal pairing.

---

[3] Following Kim, it has been taken for granted in this debate that such differential pairing between indistinguishable minds and brains is possible. There is little evidence for this is. While it is intuitive, we must be deeply skeptical about the reliability of our intuitions with respect to natural laws. Nevertheless, the possibility of indistinguishable brains under the control of indistinguishable minds illustrates a larger issue: the intrinsic properties minds possess do not appear sufficient to determining pairing, even if they should turn out to actually vary in countless subtle ways in different individuals.

[4] Kim is sympathetic with the view that spatial properties are characteristic of the physical, and hence if we ascribe minds spatial relations, we thereby concede to physicalism.

[5] A property or relation is appropriate to account for differences in causal powers if it is conceivable that the differences in causal powers could occur by virtue of the distribution of that property or relation.

Some properties and relations would not be appropriate to ground causal differences, even if minds could be distinguished on their basis. Suppose that individual pleasures had contents that were so fine-grained that no two minds could experience the exact same quantity and quality of pleasure. It would follow that no two minds could be intrinsic duplicates. Although there would be a difference between any two minds that possessed distinct causal powers, the differences between those minds would not be appropriate to account for the differences in those powers. The exact amount of pleasure experienced by a given mind is not capable

4. Interactionism is false.

I have some concerns about the metaphysical interpretation of the causal accountability principle. Despite its appeal, metaphysical laws are difficult to establish, and so we might have much stronger support for interactionism than we have for the principle. Fortunately, Kim's challenge does not depend too heavily on this interpretation. We can provide stronger support to a weaker conclusion by adopting an epistemological interpretation. Consider the plausibility version of the causal accountability principle:

> Causal Accountability Principle (epistemic version): The plausibility of a theory is proportionate to its prospects for accounting for all of the differences in the causal powers of the objects it postulates in terms of their intrinsic properties and relations.

This version of the principle does not demand *of reality* that differences in causal powers be accounted for. It only demands that *we* treat brute differences in causal powers as a theoretical cost. A theory that cannot account for minds' discriminating powers may not be false a priori, but it still deserves a lesser share of our credence.

The greater the number of unaccountable differences in causal powers, the worse for the theory. Interactionists are committed to a huge number of differences in order to distinguish the powers of every individual mind.[6] So while the plausibility version of the principle does not outright exclude the possibility of inexplicable interactions, it suggests that we will need very strong reasons to believe that all of the required interactions are inexplicable.

The revised argument goes as follows:

1. The plausibility of a theory is proportionate to the prospects of accounting for all of the differences in the causal powers of objects it postulates in terms of their intrinsic properties and relations.
2. If interactionism is true, then intrinsically identical minds can differ in their power to influence intrinsically identical brains.

---

of pairing minds and brains, for the structure of pleasure is too radically unlike any of the properties that we know to individuate brains, and too distinct from the mental operations of decision-making, to plausibly pair brains with minds.

[6] There are currently billions of human minds, trillions of vertebrate minds, and untold numbers of alien minds. There were many more minds in the past (that presumably would have controlled only their individual bodies had they managed to survive to the present) and no limit to the number of possible future minds.

While these minds and their brains are phenomenally and intentionally distinctive, if intrinsically identical minds and brains can have inexplicable differences in causal powers, it is unlikely that intrinsically distinct minds and brains possess their different causal powers in virtue of their non-causal differences.

3. If interactionism is true, then there are very many minds that bear no relationships to their brains that are appropriate for accounting for their powers.

---

4. Interactionism is implausible.

Some facts about causal powers may be brute, but, for reasons I will explain below, interactionists cannot cleanly solve their problems by positing brute powers. There is a cost to unaccountable differences in causal powers, and given the tremendous number of unaccountable differences interactionists require, their view would be deeply implausible.

The Neural Discernment Problem

In the last section, I presented the pairing problem and suggested that it can be strengthened by adopting an epistemological interpretation of the causal accountability principle. The remainder of the paper will concern the neural discernment problem. Not only must minds be capable of controlling only individual brains, but they must do so by controlling only particular neurons. Since our behavior depends upon patterns of neural activity in our brains, interactionism requires minds that excite or inhibit the firing of groups of neurons. They must be able to influence some neurons and not others, and a mind's capacity to influence must change from moment to moment.

The problem results from the similarity of the neurons that are influenced by a given mind with those that are not influenced. Marsel Mesulam (1990) summarizes the relevant neurological details:

> Neurons have a limited number of actions, that is, they either fire or they do not. And yet, individual neural networks underlie vastly different cognitive operations. Such variations in behavioral affiliation are not based on differences in the nature of the constituent neurons but on differences in the type of input, the access to output, and the architecture of intermediary processing. Behavior is not contained in the neuron or in the anatomical site but in grids of connectivity that are both localized and distributed. Such networks allow a very large number of computational operations to be associated with specific cognitive processes.

In other words, neurons' functions are primarily a matter of the connections that they bear to other neurons and not a matter of their distinctive properties. For instance, the neurons in the motor cortex that are responsible for the control of our feet do not exhibit systematic intrinsic differences with the neurons that

are responsible for the control of our hands.[7] Each neuron resembles countless others. Neurons are functionally distinguished by their connections. Minds must exert some control over particular neurons that cannot be explained by a match between their respective known intrinsic properties.

The neural discernment problem is the problem of accounting for the causal pairing between minds and neurons. An individual mind in a given mental state may produce one brain state by opening ion channels in some neurons and closing ion channels in others, but the ion channels that are opened and closed are essentially indistinguishable. Given their physiological similarity and the fact that mental states exhibit no differences remotely like the differences in physical effects they produce, it is dubious that a mind could effectively target particular groups of neurons. Interactionists owe some alternative account of how minds accomplish this feat.

The problem is exacerbated by the complexity of neural processes and the dispersal of cognitive function. In order to produce coherent behavior, a mind would need to exert very precise control over its brain's neurons. While it is possible that some neurons are capable of initiating behavioral responses all by themselves, it is deeply implausible that individual neurons are solely responsible for many complex behaviors.[8]

Even among the minority of cognitive scientists sympathetic with the existence of grandmother neurons, the possibility that complex concepts or propositions are represented by single neurons is not taken seriously. Jeffrey Bowers (2009), for instance, mounts an intriguing defense of the claim that grandmother neurons may play a role in face or word recognition, but does not go so far as to suggest that they are responsible for the representation of whole scenarios (or, presumably, complex action plans).

Controlling action requires more than exciting or inhibiting all of the neurons

[7] This is not to say that neurons are identical. On the contrary, neurons can be distinguished from each other in a variety of ways: size, arborization, mylination, and neurochemistry. However, these distinctions don't correspond with the sorts of distinctions that would be needed to guide neural discernment. It is possible that there are elusive neurons whose distinctiveness could ease discernment, but interactionists should not rest their hopes on the oversights of neuroscience.

Even if some neurons were distinguished by the structure, they would most likely be composed of the same parts as other neurons. The problem would re-arise at a lower level as the ion channel discernment problem.

[8] According to the dominant view in the neuroscience of cognitive representations:

> active representations in the mind are thought to correspond to the patterns of activation generated over a set of units... This representational scheme is held to apply to essentially all kinds of cognitive content: Words, letters, phonemes, grammatical structures; visual features, colors, structural descriptions of objects; semantic, conceptual, and schema representations; contents of working memory and contextual information affecting processing of current inputs; speech plans, motor plans, and more abstract action plans — all are thought to take the form of distributed patterns of activation over large neural populations. (Rogers and McClelland 2014, 15)

in a neural subregion. The groups of neurons responsible for the control of individual actions are not restricted to particular locations in the brain (Cisek 2007; Cisek and Kalaska 2010; Hommel and Elsner 2009; Jeannerod 2006). So minds must have some way of discerning between billions of neurons to excite only those necessary to produce their action.

Consider what is required to pick up a cup of tea: when we consciously decide to pick up the cup, we don't actively choose the degree to contract the muscles of our arms, when to start tightening our fingers, or how to adjust the speed and tilt of the cup to avoid spilling its contents. These decisions are made subconsciously unless consciously overridden.

Even if interactionism is true, surely many of the unconscious details of action guidance are not settled by minds or mental states, but are instead calculated within brains.[9] In order to perform these calculations, brains require access to information about the action that is to be performed. Since the calculations are sensitive to the intended means and outcome, the brain must represent this information in some way or other. In order to pick up a cup of tea, the mind must produce something at least as complicated[10] as a neural representation of the goal of having the cup raised and supported.

Minds might be able to excite the right neurons to construct representations of their goal and then transfer control to unconscious biological calculators to prepare the requisite motor mechanisms. In order for a mind to direct its brain to pick up a cup of tea, the mind would need to project both the goal state and an abstract representation of the means toward that goal in the patterns of neural firing in the brain.

Creating a cognitive model of the goal of an action is a complex task. The representations need to be sufficiently intricate to make room for the details

[9] One example of this comes from grasping studies:

> Whether [people] grasp objects with one orientation or another depends on how they plan to orient the objects. Where along the lengths of objects they grasp the objects depends on the height to which they plan to carry the objects... Plans can take into account a multiplicity of factors, including bio-mechanical efficiency and comfort, the relative importance of different kinds of costs such as the symmetry or asymmetry of bimanual movements, and considerations of other needs." (Rosenbaum et al. 2012, 26).

These studies indicate that our behavior is sensitive to unconsciously computations of considerations relevant to consciously set goals. It is possible that these unconscious activities are performed by the mind rather than the brain, but attributing them to the mind will only transform the burden from that of instigating a goal representation to instigating a set of motor commands, which is likely to be no less complicated.

[10] There might be no clean separation between perceptual representation, outcome valuation, and goal description in the brain (Anderson 2014). This would relieve interactionists of having to explain how a mind might produce a representation of a goal within the brain, but it doesn't solve their fundamental challenge. Minds would still need to influence the factors that go into producing action. Implanting a goal simplifies the process. Influencing action through outcome evaluation, the perception of affordances, fine motor representations, or something intermediate between these would be at least as complex.

required to override any default choices made by the brain. Given that we can choose to extend a finger or deliberately spill a little tea as the cup is raised, we are capable of configuring fairly fine-grained representational details. These details surely cannot be represented by the activity of a few, or even a few dozen, neurons.

In sum, minds require complex and precise control over their brains to produce coherent actions. Accounting for the distinct powers of minds will require interactionists to locate intrinsic properties or relations that link minds in mental states with the neural groups they influence. Contemporary interactionist theories lack the resources to provide such an account, and it is hard to see where they might find them.

We may formulate the argument as such:

1. The plausibility of a theory is proportionate to the prospects of accounting for all of the differences in the causal powers of objects it postulates in terms of their intrinsic properties and relations.
2. If interactionism is true, then minds have the power to produce different patterns of activity in distinct groups of neurons depending upon their mental state.
3. It interactionism is true, then there are many minds and many possible mental states such that there are neither differences in the intrinsic properties of minds in those states nor any variation in the relations that they have to groups of neurons that are appropriate to account for the difference in their causal powers.

---

4. Interactionism is implausible.

The neural discernment problem may seem like a minor variation of a familiar problem, but it reveals new depths to the challenge facing interactionists. In the remainder of the paper, I will argue that the most promising responses[11] offered to Kim's problem fare worse when cast as responses to the neural discernment problem.

[11] I will leave Alvin Plantinga's theistic intervention response out of the following, though I take it to be perfectly adequate for those with the right prior commitments. Plantinga (2007) observed that pairing might be achieved by divine concurrence. God only concurs with some of the possible causal outcomes of our mental states. In the context of the present proposal, while our intentions may be sufficient to trigger every neuron in our brains (or every neuron everywhere), God only concurs with a causal effect on a very specific subset of neurons

Omnipotence can overcome many philosophical problems. The chief drawback of this proposal, aside from its theistic requirements, is that it demands quite a lot more divine intervention than metaphysicians have lately been tempted to accept. If our minds are incapable of singling out the neurons that and are not to be activated, then God must be responsible for deciding and producing the behavioral products of all of our intentions.

Responses

1) Pairing with Spatial Relations

There is something deeply unintuitive about the idea that minds or their states could have spatial extensions, shapes, or relative positions, and this has led many dualists to infer that minds must be non-spatial. These intuitions have been questioned (Bailey, Rasmussen, and Horn 2011; Lycan 2009), and spatially located minds fit well with modern forms of property dualism that take mental properties to inhere in physical objects.[12] If minds occupy regions of space (or if their properties inhere directly in spatially located objects), then their (inherited) spatial relationships could account for their causal effects.

According to *pairing with spatial relations*, minds' decisions affects their brains and not others because of their spatial relations. Despite the apparently different nature of minds and material objects, no a priori considerations strictly prevent minds from being located in space or entering into spatial relations. The question of whether or not minds have spatial relations is orthogonal to the standard arguments for dualism, which depend instead on the nature of intrinsic mental properties. Spatiality should not be a defining criterion of the physical, and Kim's arguments against spatial locations for minds are open-ended and unconvincing.[13]

Attributing locations to minds would help take care of Kim's worries, since spatial relations would neatly account for the pairing between individual minds and brains. Nevertheless, this solution will not resolve the interactionists' problems unless it also provides a reasonable response to the neural discernment problem.

In order for the locations of minds to be capable of accounting for their distinctive effects, minds must be uniquely located with respect to their brains. The pairing problem would still arise if minds were omnipresent or if many minds overlapped each brain, for minds would then have identical spatial relations with the brains they could and could not influence. While no one has seriously proposed that human minds are large enough to overlap many different brains, it is highly plausible that individual minds would overlap with many different neurons.

In order for a mind's location to account for its neuronal influence, the mind must be located in a specific way with respect not just to it's brain, but to those

[12] Kim was explicitly concerned with substance dualism rather than property dualism, but the issues facing the latter view are similar. Property dualists, by attributing mental properties to physical objects, may secure some connection between physical and mental properties by virtue of co-inherence. The appeal of this route and its problems mirror those of pairing by spatial co-location discussed in this section, or pairing by union relations discussed in the next.

[13] Kim's argument takes the form of a series of difficult questions for interactionists: Where are minds located? Why don't minds count as material objects? What happens when two minds both overlap a single brain? How does the mind have enough structure to produce complex effects on the brain?

A difficult question is not an argument, but rather an appeal to intuitive implausibility. Our intuitions about such things are often malleable, and are subject to overturning as the field of conceivable theories expands. The last of Kim's difficult questions is the most serious, since it carries a hint of the neural discernment problem.

neurons it affects. The obvious extension of the proposal to handle the neural discernment problem holds that minds are extended in space and influence all and only those neurons with which they overlap. Overlap isn't the only way to distinguish mind-neuron pairings, but there are few alternatives that are nearly as natural, and they will share the same problems.

If, as neuroscience seems to indicate, the neurons that are involved in initiating a given action are not highly localized, then in order to secure the pairing between minds and (only) the right neurons by overlap, minds would need to adopt heavily contorted shapes. Furthermore, since the neurons needing influence change from moment to moment, the mind's shapes would need to be in flux.

The unintuitiveness of this picture speaks against it. There is nothing strictly impossible about minds that undergo frequent spatial contortions to occupy hyperspecific dimensions as a response to phenomenal and intentional states, but it is deeply at odds with our conception of our own minds and with our understanding of fundamental natural processes.

A separate problem arises from the reduction of a mind's causal powers to its location. On this proposal a mind's specific causal powers would be largely attributable to it's shape. The causal effect of the decision to raise one's hand has little to do with its intrinsic qualitative feel or its intentional content. Instead, it is a matter of the overlap of the mind with the neurons responsible for hand raisings.

This threatens the advantages of interactionism over epiphenomenalism. Interactionists often arrive at their position in part because of the counter-intuitiveness of epiphenomenalism. But it is doubtful that attributing a mind's causal powers to non-mental properties of a fundamentally different structure preserves the advantage over epiphenomenalism. The problem is that arbitrary causal connections are no better than no connections at all. If minds exert themselves through their shape, and the shapes that they adopt are arbitrarily correlated with their mental properties (even if they are fortunately just right to exercise rational control over a human brain), then their mental properties are causally arbitrary. Note that the shape that a mind adopts could be the result of its purely mental properties, but by producing effects through its geometric properties, the mind's mental properties are causally irrelevant qua mental properties.[14]

Consider, for instance, the view that the intrinsic characters of phenomenal

[14] This complements Chalmers (1996) response to the purported advantages of interactionism over epiphenomenalism:

> Even on [interactionism], there is a sense in which the phenomenal is irrelevant. We can always subtract the phenomenal component from the explanatory account, yielding a purely causal component... Some might argue that psychons (or ectoplasm, or whatever) are entirely constituted by their phenomenal properties. Even so, there is a sense in which their phenomenal properties are irrelevant to the explanation of behavior; it is only their relational properties that matter in the story about causal dynamics. (157-58)

properties are sufficient to bring about certain actions. Hedda Mörch (2018) suggests:

> knowing how pain feels, it is hard to see how it could make any subject do anything else than try to avoid it solely in virtue of feeling like that. Try to imagine the pain that [someone who has never felt pain] would experience from stepping on [a] nail. The quality of such an experience can only be described as intrinsically repulsive. Could this quality – alone, in and of itself and in the absence of interfering motives – make someone who experiences it try to pursue more of it, remain indifferent to it, or otherwise do anything else than try to avoid it? This seems very hard to conceive of. (2018, p. 304, emphasis in original)

This view is very attractive until we consider what it means for neural discernment, for feelings cannot produce avoidance behavior directly. The quality of an experience can only produce an appropriate behavioral response by influencing very specific neurons. On the present proposal, there must be something special about the locations of neurons such that pain is especially liable to influence neurons in those locations. The intrinsic feeling of pain has no obvious bearing on that.

Furthermore, if minds' causal powers are bound to ever-changing shapes, then there must be some explanation of their dynamics. Why should a mind that decided at one moment to raise its hand then overlap the relevant region of the premotor cortex? In other words, why should the intentional act be correlated with its specific movement? The psychological laws thought to govern mental states cannot be easily extended into laws governing the shapes of minds.

The challenge of explaining the dynamics of mind-shapes is especially pressing given that human brains are not organized in accordance with a single precise plan (let alone to a plan common with the brains of other animals). The laws that govern the dynamics of mind-shapes must be sufficiently sensitive to accommodate widespread differences in neuronal positioning. Without a reasonable explanation of minds' contortions, overlap cannot get us far.

In summary, while spatial relations might assist with the pairing problem, they provide less help with the neural discernment problem. Not only would minds have to be located in space, but they would have to be in constant flux between labyrinthian shapes to overlap with all and only the right neurons to produce coordinated behavior. It is implausible that minds take such shapes. It is also implausible that our agency could be preserved within a framework in which the contours of our minds played such an integral role in action guidance. So pairing with spatial relations is not a promising avenue of response for interactionists.

2) Pairing with Union Relations

Although Descartes argued that minds must be non-spatial he was also acutely sensitive to the problems surrounding mind-brain interaction. When discussing the capacity of minds to influence individual brains, he repeatedly made reference to a union between them. According to *pairing with union relations*, minds' unions with their brains account for causal pairing.[15]

Kim criticizes this proposal for its explanatory shallowness. According to Kim (2005), "the word 'united' merely gives a name to the mystery rather than clarifying it" (78). While Kim might agree that we can provide a formal outline of the shape that a pairing relation would need to take, he suggests that without the details, it is of no help to interactionists.

In his letters with Elizabeth, Descartes admitted that the notion was difficult to square with the distinctiveness of mind and body, so we can grant Kim that Descartes had nothing more than the form of a solution: pairing with union relations amounts to the proposal that there exists some relation or other that ties our minds with our bodies, and that differences in that relation can explain differences in the causal powers of minds.

Understood in this way, the proposal can help resolve the pairing problem. We may not understand union relations, but our imaginative failure is just that.[16] Scientists routinely postulate unknown and unfamiliar phenomena to plug explanatory holes, and it might reasonably be claimed that some of the most basic aspects of reality – laws of nature, objective chances, norms, even spatial and temporal relations – remain dimly understood. This is no reason to deny such postulates roles in our theories, so the shallowness of our grasp on unions shouldn't preclude us from putting them to work.

A direct relation between minds and brains could explain the discerning causal influence of minds in much the same way as could spatial overlap. Despite their importance to many ordinary causal pairings, there is nothing that makes spatial relations especially well-suited for pairing. Although we lack clear examples of union relationships participating in other causal pairings, the possibility is coherent.

Unions offer less help with the neural discernment problem. The difficulties arise from the complexity required to underlie neural discernment. Minds could not simply be united with their brains as a whole; their union(s) would have to account for the effect they have on distinct groups of neurons. Minds must either have a number of unions with different parts of the brain (through their mental states) or else they must have a union relation that is sufficiently complex to produce many different mental effects when the mind is in different states.

[15] Foster (1968) proposes that we need a primitive notion of 'ownership' to make sense of mind-brain pairing. Whether or not it is a proper instance of the present proposal, it bears a sufficient resemblance to share its problems.

[16] David Jehle (2006) makes a similar point.

Both versions of the proposal face the same challenges. First, these views are committed to highly complex postulates that count against the parsimony of the theory, either in the form of a large number of distinct unions or in a single multifaceted union. In order to maintain a sophisticated behavioral repertoire, minds must exhibit different relationships to different groups of neurons.

Second, and more significantly, these unions would need to be established at some point in the development of the brain. It is not obvious how the creation of novel unions of the right sort could figure into the process of neural development. Neurons are no more distinguishable in the developing brain than they are in the mature brain, so the forces responsible for establishing particular union relations between minds and neural groups would be susceptible to their own discernment problem. It is no help to say that minds are united with just the neurons that they need to affect if we cannot begin to explain how they got that way.

Union are coherent, but they are not well suited to the specific work necessary to pair minds with groups of neurons. The necessary unions would introduce a great degree of complexity and would need explanations themselves.

3) Pairing with Intrinsic Properties

If we cannot solve the neural discernment problem with relations, it may be possible to do so with intrinsic properties.[17] According to this idea, pairing does not depend on the relations between minds and brains, but on how their properties align. Minds have the power to affect all and only those brains whose intrinsic properties complement their own. Nothing constrains the capacity of minds to influence brains except for the rarity of their complements. This is the *pairing with intrinsic properties* proposal.

The chief challenge to this proposal is finding complementary properties up to the task. The proposal places demands on both minds and brains: on the assumption that known mental properties can't discern brains by their known physical properties, minds and brains must have hitherto undiscovered properties that make them uniquely suited to mutual influence.

In order to explain the incapacity of minds to control multiple brains, individual minds and brains must align in highly unusual ways. It is implausible that our minds should just have happened by simple chance to have properties that complement the right brains. The *kindling hypothesis* explains this by suggesting that brains give rise to minds with just the right intrinsic properties to subsequently influence them.[18] Brains have properties that make them

[17] David Jehle (2006) and Brad Saad (2017) develop versions of this response.

[18] Bailey, Rasmussen, and Horn (2011) briefly consider a version of this hypothesis. They attribute the persistence of a mind's causal powers to its historical properties: the fact that the mind was created by the brain is sufficient to explain why the mind continues to hold sway over that brain. It is more plausible that a brain might produce a mind with the standing properties necessary to subsequently influence it, because purely historical properties are not known to play roles in causation.

amenable to certain kinds of influence, and those properties are partly responsible for producing complementary properties in the generated mind. The initial connection between mind and brain requires no complementarity, and it explains how minds and brains subsequently become so fortuitously suited to influencing one another.

As with the previous proposals, pairing with intrinsic properties encounters special difficulties with the neural discernment problem.

First, there is the issue of parsimony. In order for pairing with intrinsic properties to solve the pairing problem, there must be complementary properties belonging separately to minds and their brains. This only requires one complimentarity per pair: a large number, but perhaps not outrageously so. However, in order to accommodate the mind's discerning influence over different groups of neurons, the properties of the mind could not merely complement the properties of its brain, they would have to complement each of many different groups of neurons in different ways. Since each different action the mind is capable of instigating will require a separate neural group, and since a very large number of neural groups would be needed to account for our behavioral flexibility, interaction would require a tremendous number of distinct individually complementary properties. Even if we allow for the possibility of mental properties, nothing we know about the natural world suggests that physics should play along. This makes the existence of such complementary properties look like a miracle.

Second, there is the matter of development. Each group of neurons must acquire the right intrinsic properties to be influenced by its mind sometime during the process of development. If the relevant groups of neurons are only distinguishable by their eventual interconnects, there is no way for them to be singled out to acquire the right intrinsic properties in the first place.

The kindling hypothesis goes some way towards explaining how minds and brains might pair, but it does not explain how minds and neurons do. It helps insofar as the kindling process does not need to be discerning: brains don't have to identify an appropriate mind to pair with if they create a mind with the needed properties. The complementarities necessary to explain neural discernment, being far greater in number, cannot all be established with a single simple act. The alignment between different mental states and tens or hundreds of thousands of groups of neurons can't be kindled, so their special fit with each other demands a better explanation.

Pairing with intrinsic properties ultimately suffers from similar problems as the previous proposals. It may be suitable for explaining how individual minds are causally paired with individual brains, but individual minds must also be causally paired with a huge variety of different neural groups. The maneuvers required to accommodate such variety detracts from the theory's parsimony and introduce issues in development that mirror the original problem.

4) Brute Mind-Brain Causation

According to the first two proposals, relations between minds and brains account for their causal pairing. According to third proposal, pairing is achieved through the matching of intrinsic properties. These proposals all concede that some explanation of pairing is needed. In contrast, several commentators[19] have responded to the pairing problem with skepticism about the causal accountability principle. They object that there could be differences in causal relations that do not coincide with qualitative or relational differences.[20] Some physical phenomena might be like this (Tooley 1990). Though it is somewhat less plausible that causal powers, rather than causal relations, are brute, we may have no principled reason to permit only the latter.

If we surrender the causal accountability principle, we may attribute unaccountable differences in causal powers. According to *brute mind-brain causation*, there is no accounting for the capacity of minds to influence some brains and not others.

The plausibility version of the causal accountability principle permits brute causation, but it comes with a cost. Even if brute causation is possible, we have good reason to go this route only as a matter of last resort, for brute causal powers would be deeply mysterious.

By embracing brute causal powers, we relinquish the explanatory resources we might use to make sense of their patterns. If it is true that minds exert brute influences over particular brains, then there is no reason why a given mind should influence the same brain at each moment. A mind could just as easily routinely switch – perhaps controlling one brain one day and another brain the next.

Since independent brute relationships have no reason to follow any patterns, bruteness essentially deprives us of the resources for explaining consistency. The fact that causal relations are well-ordered is strong evidence that they are explicable. If they are explicable, then we need sufficient resources to explain them.

Brute causal relations are possible, but brute mind-brain causation is an unsatisfying answer to the pairing problem given the consistency of mind-brain influences, even leaving aside the additional considerations introduced by neural

[19] See Audi (2011), Bailey, Rasmussen, and Horn (2011), and Plantinga (2007). Kim introduces the pairing problem in the context of a discussion of reductive theories of causation. The viability of brute causation will depend in part on the viability of reductive theories: Humean approaches to causation cannot make sense of brute causation. If reduction about causation is required to motivate the pairing problem, then the argument will be dialectically ineffective, because dualism correlates with opposition to such theories.

[20] There is a difference between brute causality, in which there is no explanation of the causal relation, and haecceitistic causation, where causation is explained by haecceitistic differences. This distinction is important in part because haecceities may persist through time and thereby explain diachronic patterns of interaction. Haecceistitic causation falls under the third proposal.

discernment. Nevertheless, the neural discernment problem exacerbates the interactionists' difficulties.

The intricate causal control necessary for intelligent behavior necessitates complex causal regularities. It would not merely be a brute fact that a mind influences one brain rather than another, it would need to be a brute fact that a mind could separately influence thousands of physiologically similar but functionally distinct neurons one moment and thousands of other neurons in the next. By making the causal relation between minds and their brains' neurons brute, we abandon any explanation for why our mental states influence one group rather than another. Our decisions may still be causally responsible for our actions, but there is no reason why our mental states should produce the actions that they do.

Brute mind-brain causation also inherits some of the problems of epiphenomonalism.[21] While brute causal interactionism would not rob our minds of the power to affect things in the world, it would prevent us from explaining how our behavior results from our mental states. If there is no reason why our mental states affect the particular groups of neurons that they do, then their powers are arbitrary. There can be no explanation of why a decision to raise an arm leads to raising an arm rather than stomping a foot. Our mental states would not explain our behavior, even if they would cause it. This is clearly problematic, and removes much of the appeal of interactionism over epiphenomenalism.

Conclusion

While the neural discernment problem is similar to the pairing problem, it offers additional challenges to interactionists. The most promising answers to the pairing problem come at a theoretical cost; they either complicate dualistic theories by postulating ad hoc explanatory mechanisms or they try to make due without explanations. For the pairing problem, the cost of these proposals is low enough to be worth paying. The costs of applying these same responses to the neural discernment problem are much steeper, so even the most exuberant interactionists should hesitate to pay. There may yet be some theoretically affordable way to explain our minds' abilities to discern which neurons to influence. Until this explanation is produced, interactionism will remain implausible.

Bibliography

Anderson, Michael L. 2014. *After Phrenology: Neural Reuse and the Interactive Brain*. MIT Press.

Audi, Paul. 2011. "Primitive Causal Relations and the Pairing Problem." *Ratio* 24 (1): 1–16.

[21] This may be inevitable. See footnote 14.

Bailey, Andrew M., Joshua Rasmussen, and Luke van Horn. 2011. "No Pairing Problem." *Philosophical Studies* 154 (3): 349–360.

Bowers, Jeffrey S. 2009. "On the Biological Plausibility of Grandmother Cells: Implications for Neural Network Theories in Psychology and Neuroscience." *Psychological Review* 116 (1): 220–251.

Chalmers, David J. 1996. *The Conscious Mind: in Search of a Fundamental Theory*. Oxford University Press.

Cisek, Paul. 2007. "Cortical Mechanisms of Action Selection: the Affordance Competition Hypothesis." *Philosophical Transactions of the Royal Society B: Biological Sciences* 362 (1485): 1585–1599.

Cisek, Paul, and John F Kalaska. 2010. "Neural Mechanisms for Interacting with a World Full of Action Choices." *Annu. Rev. Neurosci* 33: 269–98.

Foster, John A. 1968. "Psychophysical Causal Relations." *American Philosophical Quarterly* 5: 64–70.

Hommel, Bernhard, and Birgit Elsner. 2009. "Acquisition, Representation, and Control of Action." *Oxford Handbook of Human Action*: 371–398.

Jeannerod, Marc. 2006. *Motor Cognition: What Actions Tell the Self*. Oxford University Press.

Jehle, David. 2006. "Kim Against Dualism." *Philosophical Studies* 130 (3): 565–78.

Kim, Jaegwon. 2001. "Lonely Souls: Causality and Substance Dualism." In *Soul, Body, and Survival*, edited by Kevin J. Corcoran. Ithaca: Cornell University Press.

———. 2005. *Physicalism, or Something Near Enough*. Princeton University Press.

Lycan, William. 2009. "Giving Dualism Its Due." *Australasian Journal of Philosophy* 87 (4): 551–563.

Mesulam, Marsel. 1990. "Large-Scale Neurocognitive Networks and Distributed Processing for Attention, Language, and Memory." *Annals of Neurology* 28 (5): 597–613.

Mörch, Hedda Hassel. 2018. "The Evolutionary Argument for Phenomenal Powers." *Philosophical Perspectives*: 293–316.

Plantinga, Alvin. 2007. "Materialism and Christian Belief." In *Persons: Human and Divine*, edited by Peter van Inwagen and Dean Zimmerman, 99–141. Oxford University Press.

Rogers, Timothy T, and James L McClelland. 2014. "Parallel Distributed Processing at 25: Further Explorations in the Microstructure of Cognition." *Cognitive Science* 38 (6): 1024–1077.

Rosenbaum, David A, Kate M Chapman, Matthias Weigelt, Daniel J Weiss, and Robrecht van der Wel. 2012. "Cognition, Action, and Object Manipulation." *Psychological Bulletin* 138 (5): 924.

Saad, Bradford. 2017. "Interactionism, Haecceities, and the Pairing Argument." *Inquiry* 60 (7): 724–741.

———. forthcoming. "Indeterministic Causation and Two Patches for the Pairing Argument." *Pacific Philosophical Quarterly*.

Tooley, Michael. 1990. "The Nature of Causation: a Singularist Account." *Canadian Journal of Philosophy* 16: 271–322.