

Peter Ulric Tse

The Neural Basis of Free Will: Criterial Causation.

Cambridge, MA: MIT Press 2013.

432 pages

\$38.00 (cloth ISBN 978–026201910–1)

Neuroscientist Peter Tse aims to solve two longstanding philosophical problems in this book: the problem of free will and the problem of mental causation. He aims to do this by presenting an account, which is somewhat speculative but which is solidly grounded in contemporary neuroscience, of how the brain processes information. This account is a significant achievement and will be of great interest to philosophers of cognitive science. Oddly, however, there is little in the book to interest philosophers whose main focus is on the problems that Tse aims to solve. Neither his account of mental causation nor that of free will – insofar as he accounts for either – addresses their concerns. Rather, the picture Tse presents is neither novel nor illuminating insofar as it is relevant to these philosophical concerns. Far from solving the problem of mental causation, Tse presents an essentially orthodox picture of information processing in the brain, a picture that provides the background against which the problem arises, rather than solving it. He does venture some speculations about how the problem can be solved, but the best of these suggestions are not novel nor are they developed in any depth, while many are irrelevant or confused. Similarly, though he presents us with a model of indeterministic decision-making in the brain to which libertarians might want to help themselves, this picture does not aid them in showing how indeterministic information processing might be free.

Tse believes that the problem of mental causation and the problem of free will are intimately related. Indeed, they have a common solution: criterial causation. Criterial causation is causation by physical events, but as realized in the brain such causation is causation by information. Neurons and neuronal circuits fire in response not to energy, but to *patterns* of energy: they fire when the patterns of energy satisfy a sufficient number of the criteria to which the neurons are attuned. Because neurons respond to patterns of energy, which are informational, eliminativism is false: cognition is caused by meaningful events. Further, the downstream effects of criterial causation include behavior and the fine-tuning of the very criteria to which neurons will, in the future, respond, and the sculpting of new neuronal circuits. Since this is causation by the mind, criterial causation is the solution to the problem of mental causation. Criterial causation is also the solution to the problem of free will. In altering the criteria to which neurons will respond, criterial causation sculpts future behavior. And that is free will. If it happens indeterministically, which Tse thinks is probable, it is *strong* free will: we genuinely possess alternative possibilities and the power to choose between them.

The account of criterial causation in the brain, described in great detail and with far more attention to lower-level properties of neurons than is common in literature aiming to integrate philosophy and neuroscience, is the portion of the book which philosophers interested in cognitive neuroscience will find most valuable. Tse argues that neurons have properties that allow for a rapid resculpting of the criteria to which they respond, properties that allow them to soft-wire circuits of neurons to respond to ever-changing informational properties. But it is hard to see how any of this amounts to a solution to the problems of mental causation and free will.

Though many of the details are novel and interesting, so far as these problems are concerned, Tse presents a rather orthodox picture of cognition, not a million miles from the constraint satisfaction picture of connectionism. Neural circuits are physical circuits, set up in such a way as to respond preferentially to some (equally physical) inputs and not others. That's common ground for all players in the debate over mental causation: giving us a novel account of how it is done isn't solving the problem of mental causation, it is providing the background which the problem assumes. The problem of mental causation is not the problem of explaining how physical processes can carry information, as Tse seems to think. It is the problem of explaining how mental properties can be causally relevant *qua* mental properties. It requires, that is, an account of what role mental properties might play, given that physical processes seem sufficient for all the explanatory work. Solving it does not require an account of how cognition is realized in the brain: it requires a metaphysics of mind.

That is not, of course, to say that the account of how cognition is realized in the brain is irrelevant to the problem of mental causation. The account might make some solutions more attractive and others less so. For instance, an account of how cognition works might motivate an identity theory, and the identity theory might offer a solution to the problem of mental causation (by asserting that causation by neurons or neuronal circuits *just is* mental causation). There are, of course, well known objections to identity solutions to the mind-body problem, as well as arguments against the claim that the identity theory solves the problem of mental causation. A different set of problems confronts functionalist accounts of mental causation. The theory Tse offers does not clearly lend itself to any of these views (it might be developed in ways consistent with more than one, *pace* some ill-judged remarks about functionalism). Whether Tse's account might help to motivate a genuinely satisfactory account of mental causation is a question we cannot answer until the hard work of developing and defending such an account is performed.

Why does Tse think he has an account of mental causation, rather than an account of cognition, which provides the basis for the problem? Tse makes several different claims as to why criterial causation is mental causation. Several of these turn on the further claim that the brain operates indeterministically, and that therefore physical processes might take any of several or many routes. Only some of these routes are informational, because only some configurations of particles are the kinds of patterns our brains are set up to detect; criterial causation might therefore be seen as selecting from among those physical processes only the ones that are mental. This, Tse claims, shows that mental causation is not mere physical causation. But of course he is also committed to claiming that the process is realized physically, that is, through the physical properties of neurons. Once again, all parties agree that only some physical chains carry information and, furthermore, that the brain is set up to respond to some physical processes and not others. The fact that a physical event causes brain processes only if it meets certain criteria does not entail that the physical event is not physical or that it causes events *qua* mental. Again, the mere physical event seems sufficient to explain the observed effect, leaving the problem untouched.

Tse has another argument for why criterial causation is mental causation, but it is badly confused. It rests on a misunderstanding regarding the causal closure principle. Tse understands the principle to claim that physical causes are sufficient for the occurrence of physical effects. If indeterminism is true, then physical causes sometimes or often are not sufficient for the

occurrence of later events. Tse therefore concludes that the closure principle is false for indeterministic systems, so it is no obstacle to mental causation. But the causal closure principle is, roughly, the principle that physical events can be accounted for by physical causes, or (equivalently) that physics is causally complete. It is silent on whether physics is deterministic or not. The brain may be indeterministic; causal closure remains an obstacle to mental causation.

Tse's account of free will is no more illuminating. Much of his argument is addressed to Galen Strawson's impossibility argument, which he takes, rather oddly, to be the central argument against the existence of free will. Strawson's argument, he claims, rests on the impossibility of self-causation. But criterial causation entails the very rapid resculpting of neural circuits: not self-causation, but causation of the criteria to which neurons will respond. So we don't need self-causation at all. Of course that's not answering Strawson's objection: it is refusing to engage with it. Strawson's argument is a regress argument, turning on the claim that even though it is true that we do not simply respond as nature/nurture has formed us, because we change the way we respond, these changes are themselves explained by nature/nurture. Self-causation is mentioned only as a possible regress stopper. Saying that criterial causation sculpts how we respond is not solving Strawson's problem; it is giving details for how a process that Strawson describes only abstractly actually occurs. Tse's account, or something broadly like it, is taken for granted by Strawson in setting up the problem: it is no solution to a problem to note that the conditions giving rise to it obtain.

Of course Strawson's argument plays a relatively small role in debates over free will. More central is the concern over whether agents can exercise freedom-level control over their actions if how they act is sometimes undetermined. Tse promises to show how indeterminism entails not merely metaphysical openness but also control by us over which alternative is realized. It is fair to say this promise is not fulfilled. Rather, Tse is silent on the standard objections turning on luck. Again, he provides the details for an account the broad outline of which libertarian philosophers have supposed to be true. The major criticism addressed to these accounts has never been that they are empirically false (philosophers are not scientists). Rather, the major criticisms have for the sake of argument assumed that they are true, and held that they are incompatible with freedom-level control. Tse does not show how his account avoids these criticisms: he does not even seem to be aware of their existence.

I do not want to minimize Tse's genuine accomplishment in this book. The chapters on the properties of neurons and how these properties may realize criterial causation may indeed constitute an important contribution to debates within neuroscience, and philosophers with an interest in these kinds of questions should certainly read and engage with it. It is unfortunate that the valuable parts of this book are embedded in a discussion of philosophical issues regarding which Tse has nothing useful and little genuinely relevant to say.

Neil Levy
University of Oxford