

Burge on Mental Causation

MARKO DELIĆ
University of Split, Split, Croatia

The article discusses Tyler Burge's views concerning the debate about the causal efficacy of mental properties, as found in his article "Mind-Body Causation and Explanatory Practice." Burge argues that a proper understanding of kind-individuation and causal explanation in science gives strong prima facie reasons for believing that mental and physical properties are not mutually exclusive. He does so by analysing the strength of two metaphysical theses which standardly underlie the debate—token physicalism and the "Completeness of physics." I present his analysis and argue that without an account of mental causation, his analysis does not support the conclusion that mental and physical properties are not mutually exclusive. Also, I question the methodological adequacy of Burge's analysis for scientific practice.

Keywords: Burge, mental causation, psychology, neuroscience, causation, explanation, overdetermination, properties, kinds.

1. Introduction

For several decades now philosophers of mind have been struggling with the question of whether, and how could mental states (and events), in virtue of their mental properties (such as their intentional or phenomenal properties) exert any causal influence on the world. The worry that they do not exert any such influence is suggested by two independently plausible metaphysical theses. First of these, the "Completeness of physics" (CP) states that *all physical effects are fully determined by law by a purely physical prior history* (Papineau 2000: 179). According to the second thesis, the "Irreducibility of the Mental" (IM) mental properties are not reducible to physical properties (Putnam, Fodor). Now, if every physical event (a certain behaviour) is completely determined by prior physical events (states of the body or the central nervous system), and the mental properties (a certain intentional content) of those events cannot be reduced to their physical properties, it seems

that mental properties are excluded from being a possible cause of that piece of behaviour.¹

The problem has come to be known as the “Exclusion Problem” (EP), most famously associated with the work of Jaegwon Kim. EP presents itself as a problem for the “non-reductivist” types of physicalism. While discussing the causal efficacy of the mental, Tyler Burge develops his views by analysing the so-called “Token-Identity Physicalism” (TIP). Being a non-reductivist position, TIP can generate the EP. In his article “Mind-Body Causation and Explanatory Practice” Burge views the “Mental Causation Debate”² (MCD) as a result of metaphysical theses which, contrary to the conviction of many contemporary philosophers, do not possess a justified theoretical motivation in the actual sciences of cognition. He proposes shifting MCD from the terrain of metaphysics to the terrain of actual cognitive sciences, whose practice Burge considers to be the main umpire in MCD. Burge’s strategy is twofold. First, Burge tries to undermine TIP on standard externalist intuitions, thus discouraging the motivation for the identification of mental and physical events. Second, he tries to bring the notion of causal powers closer to the epistemic endeavours of actual scientific practice and argue that the indispensability of mental vocabulary in psychological explanation warrants our belief in the efficacy of mental properties and makes MCD redundant.

The first section of the paper will present the EP, and TIP as understood by Burge. The second section will consider Burge’s attempts to undermine the theoretical motivation of TIP. The third section will be concerned with Burge’s understanding of causal powers and the strength of CP as a premise in EP. Finally, the last section will try to show that, contra Burge, the metaphysical debate about the causal efficacy of the mental (MCD) has real motivation in the actual practice of cognitive sciences. Also, I will argue that Burge’s understanding of the “autonomous coexistence” of psychological and neuroscientific explanation cannot be defended without an account of mental causation and giving an adequate account of mental causation forces Burge back into the very thing he aims to undermine—the MCD.

¹ The present article is based on a talk given at the annual conference of the *Society for the Advancement of Philosophy* “New topics in Philosophy” (Zagreb, 2018). Also, I would like to thank Dunja Jutronic, Dario Škarica and Ljudevit Hanžek for commentary and support in writing this article.

² The phrase Mental Causation Debate (from now on MCD) is due to Tim Crane (Crane 1995). I will use it to refer to the general debate concerning the causal efficacy of the mental.

2. *The exclusion problem*

Token identity physicalism advocates a weaker type of identity between the mental and the physical. It can be described as consisting of the next two theses:³

- 1) For every mental event x there exists a physical event y such that $x=y$
- 2) Mental properties M of x cannot be reduced to physical properties P of y (IM)

The first theses makes TIP a physicalist position by giving ontological priority to the physical (Crane 2003) while the second thesis makes TIP a non-reductive version of physicalism. The second thesis (IM) is suggested by independently plausible theses such as the *multiple realization thesis* (Putnam 1967; Fodor 1974) according to which a single mental type can be realized by different physical types; or the *explanatory gap* which exists between first-person phenomenal descriptions and physical descriptions of those experiences (Nagel 1974; Levine 1983). One way for the exclusion problem to arise is to adopt a view of causation which treats events as being causally efficacious not in virtue of them simply being events but in virtue of the properties these events possess (Crane 2003).⁴ Then, if the physical properties of a certain event, a piece of behaviour for example, are caused exclusively by the physical properties of prior events, as CP suggests, it follows that the mental properties of those prior events are excluded from being possible causes of that piece of behaviour.⁵

Burge's way of handling EP is to undermine the metaphysical basis which generates it, thus discouraging MCD. This is expressed in the *motto* at the beginning of his paper:

Materialist metaphysics has been given more weight than it deserves. Reflection on explanatory practice has been given too little. (Burge 2007: 344)

3. *Undermining token-identity physicalism*

Burge's first step is to show that the motivation for identification of mental and physical events is not justified. For Burge, that amounts to showing that the respective sciences, psychology as the study of the mental and neuroscience as the study of the nervous system, individuate their kinds in an essentially different way. While neuroscience in-

³ Only first of these is necessary for token physicalism. In conjunction with the second it becomes a non-reductive version of physicalism.

⁴ This is also how Burge sets the exclusion problem in one version (Burge 2006: 346).

⁵ Burge notices that this way of understanding CP leaves open the possibility that, although mental properties cannot be the causes of physical properties, they could nevertheless cause other mental properties. However, this would, as he immediately notices, severely limit the causal efficacy of the mental. After all, mental properties are invoked to explain behaviour.

dividuates kinds *narrowly*, only with respect to the intrinsic, bodily properties, psychological kinds are individuated with reference to their intentional content, which is relational, that is, environmentally dependent. This view is supported by well-known externalist thought experiments. Here, I will use Burge's "aluminium-twalum" example. We imagine a person A whose environment contains the metal aluminium. When A interacts or thinks about aluminium his thoughts are about the aluminium present in his environment. Now, imagine a person B who is physically identical to person A, only whose environment contains a different metal, *twalum*. Twalum is (phenomenologically and practically) indistinguishable (to both A and B) from aluminium, yet it is a metal of a different chemical kind. What seems to follow is that the respective contents of A's and B's mental states is different. While A's thoughts are about aluminium, B's thoughts are about twalum, despite the physical states of A and B being identical (Burge 2007: 316–317).

What is suggested by the thought experiment is that a certain physical event-token, which is a plausible candidate for the identification with a mental event-token, can have different intentional contents on different instantiations. This, according to Burge, is enough to show that mental events cannot be identical to physical events (Burge 2007: 350–351).

Since externalism *per se* is not the topic of this article, we will not concern ourselves with the objections raised against it here. However, in his paper, Burge discusses an objection made to his argument by Donald Davidson. Since Burge's reply reveals his understanding of kind-individuation, it will be illuminating to present it here.

Davidson argues that broad identification of kinds doesn't refute token-identity. He gives the example of a sunburn. While sunburns are individuated environmentally (with reference to an ultraviolet radiating object, most commonly the Sun), it is still plausible to identify every token of a sunburn with a certain physiologically specified state of the skin. Burge agrees with Davidson but rejects his analogy of sunburns and mental events on epistemic grounds. The difference is, Burge argues, that sunburns can be identified in purely physiological terms, without any reference to a potential environmental cause, while mental kinds do not admit such identifications. This in turn, is grounded in the fact that physiological descriptions of sunburns provide *systematic* and *explanatory* ways of individuating sunburns, while descriptions of brain states upon which mental kinds (plausibly) depend do not (Burge 2007: 352–353).

The system of intentional content attribution is the fundamental means of identifying intentional mental states and events in psychological explanation and in our self-attributions. In fact, we have no other systematic way of identifying such states and events. (Burge 2007: 354)

Physiological properties, unlike intentional properties, do not allow for a systematic and explanatory way of individuating mental kinds.

Therefore, Burge believes it to be theoretically unmotivating to insist on TIP.

4. *Questioning the role of CP*

Similar considerations underlie Burge's analysis of the role CP has in EP. In section 1, CP was presented in terms of event properties. The physical properties of a certain event are completely determined (or have their probabilities completely determined) by the physical properties of prior events. This is the same as saying that an event can cause physical effects only in virtue of its physical properties. To see how Burge understands CP and its strength as a premise in EP, we need to explain how Burge understands causation. Burge believes the best way to understand causation is to see how causal explanation works in actual science. He discusses the notion of *causal powers*. For Burge, the causal powers of an event are determined by the properties which are relevant in describing the patterns of causation in which the kind of that event enters into. Since these patterns are different, depending on the explanatory aims of specific sciences, the properties relevant for describing them will differ too (Burge 2007: 346–347).

Applying the analysis to our present subject, CP amounts to claiming that the patterns of causation identified in physical explanation need to invoke only (and exclusively) physical properties of a certain event. If, however, the patterns of causation are different, as they are in psychology, the restriction that only physical properties are relevant in determining the causal powers of an event becomes unwarranted, since the properties (and thus the causal powers of the event) needed to explain these patterns, change. The only way to insist on such a restriction would be to show that mentalistic (psychological) discourse, that is, the patterns of causation psychology describes, is either non-descriptive or non-causal, which is, as Burge points out, far from being the case (Burge 2007: 347). If the battle is fought on the grounds of actual scientific practices, as Burge believes it ought to be, mental properties are easily defensible.

If physical events have mental properties, one is not entitled to the view that only physical properties (properties specified in the physical sciences or in ordinary physicalistic discourse) determine all the causal powers of a physical event (*as opposed to merely all the causal powers associated with physicalistic explanations of the physical event*), unless one can show that mentalistic explanation is either non-causal or fails to describe patterns of causal properties. For the causal powers of a physical event that is mental might include possible effects that are specified in mentalistic explanation. No one has shown that mentalistic explanation is either non-causal or non-descriptive. Nor is either view plausible. (Burge 2007: 347)

What this shows is that only if the effect is specified as physical (as belonging to the patterns of causation described by physics) mental properties are excluded. If it is specified as mental (as a piece of inten-

tional behaviour, for example), the mental properties easily find their way back into the adequate explanation.

One problem that can be brought against such an analysis is that it is not certain whether it respects the so-called *No-overdetermination principle*,⁶ which is commonly invoked in discussions about mental causation. The principle states that physical effects cannot be overdetermined by physical and mental causes. A certain effect is said to be metaphysically overdetermined if it has two or more sufficient, but metaphysically independent (Loewer 2015: 51). A common example of overdetermination is the case when there are two shooters, each of whom kills a victim. Overdetermination is only possible if the two cause are metaphysically independent. In the case of mental events, however, the mental and physical properties are not metaphysically independent (Loewer 2015: 51)—hence the name of the principle—*No-overdetermination*. Assuming physicalism, the dependence relation is usually described as that of *supervenience* of the mental properties on physical ones, or *realization*⁷ of mental properties by physical properties. Given this assumption, the two types of properties are again in the state of competition for a certain effect. And given CP, the mental properties come out as inefficacious.

Burge rejects this argument by arguing that the view of causation it presupposes simply begs the question against the efficacy of mental properties. The view he has in mind, I believe, is similar to an account of causation which views causation as a matter of “transference of quantities” such as energy and momentum (from now TQ).⁸ Burge finds such an account adequate for physical explanation, but problematic as a model for causal explanations as found in psychology:

Why should mental causes alter or interfere with the physical system if they do not materially consist in physical processes? Thinking that they must, surely depends on thinking of mental causes on a physical model—as providing an extra ‘bump’ on the effect. The idea seems to be that a cause must transfer a bit of energy or exert a force on the effect. (Burge 2007: 358)

If causation is understood as transference of quantities, mental causes are surely excluded, since, as Burge argues, they do not materially consist in such processes. Presupposing TQ thus begs the question. What is needed to motivate CP in the kind of way which excludes the mental, Burge believes, is to show that a physical model of causation (such as TQ) is appropriate for psychological explanation. Only then would one be in position to claim that the mental and physical somehow interfere or overdetermine a given effect. However, he finds no support for such a view:

⁶ I take the term from Heil and Robb (2019), although my construal of the principle here differs from the one they give.

⁷ For supervenience, see Kim (1993). For realization Polger and Shapiro (2016).

⁸ See Dowe (2000) for a discussion of such theories of causation.

But whether the physical model of mental causation is appropriate is, again, part of what is at issue. As we have seen, one can specify various ways in which mental causes ‘make a difference’ which do not conflict with physical explanations. The differences they make are specified by psychological causal explanations, and by counterfactuals associated with these explanations. Such ‘differences’ made by psychological causes do not require that gaps be left in physical chains of causation. They do not seem to depend on any specific assumptions at all about the physical events underlying the mental causes. (Burge 2007: 358–359)

If the causal explanations found in physics and psychology do not interfere, Burge concludes, there is no reason to believe that the properties these explanations invoke are mutually exclusive, as the principle of *No-overdetermination* suggests.⁹

5. *A weakness in Burge’s analysis*

Although being fairly sympathetic with Burge’s way of analysing the role CP has in EP, in this section I will try to express my worries with his analysis. I will try to show that his analysis does not support the conclusion that mental and physical properties are not mutually exclusive and thus cannot serve Burge’s intention to present MCD as a misguided discussion.

As seen in the previous section, the way Burge tries to dispense with MCD is to show that causal explanation in psychology is not incompatible with a view of causation associated with natural sciences, thus showing that the worries surrounding the *No-overdetermination* principle are badly grounded. He concludes:

The upshot of this reasoning is that we have no ground for assuming that the failure of mental causes to interfere in the physical chain of events must be explained in terms of mental causes’ consisting in physical events. Interference would be surprising, given antecedent assumptions about mental and physical explanation. So non-interference is in no need of explanation in ontological terms. (Burge 2007: 359)

What Burge seems to be claiming is that overdetermination is problematic only if causation is understood on a model such as TQ. Only then the mental and physical somehow interfere or overdetermine a physical effect. But, as Burge argues, since such a model of causation is not appropriate for psychological explanation, one cannot presup-

⁹ In his article, Burge makes no reference to TQ specifically. Here I use it since it is suggested by the quoted passage of Burge’s article (Burge 2007: 358). Whether this account is appropriate as a model for explanations of physical events is questionable (see Dowe 2000 and Loewer 2015: 54). Furthermore, since the relationship between psychology and neuroscience is what is at issue here, Burge could have picked interventionism (Woodward 2003) which is popular among mechanistically-oriented philosophers of cognitive science. Nothing, however, depends on what account is taken as a model for physical causation, neither for Burge’s argument, nor for my objection in the next section. Here, TQ can be understood, somewhat broadly, as standing for any account of causation which will be shown as appropriate for explanations of physical events (in natural sciences).

pose that it excludes mental causation. The only way one could show that TQ excludes the mental would be to show that that explanations in natural sciences and psychology interfere and this is not the case. I agree with Burge on this point. However, I do not agree with the jump Burge makes from the premise that physical and psychological explanations do not interfere to the conclusion that the properties these explanations invoke are not mutually exclusive. To show that this is the case, Burge would have to give an account of causation which would be *both appropriate for psychological explanation and compatible with a model of causation such as TQ*. And such an account is exactly what is missing in Burge's analysis. Without an account of mental causation, I see no justifiable reason for Burge to uphold such a "compatibility" between different types of causation. What I am claiming is that the burden of argument is on Burge. Although he is correct in claiming that TQ is not sufficient to exclude mental causation, he is not also justified in claiming that TQ is compatible with mental causation. Such compatibility requires a positive argument and Burge does not give one. Burge tries to argue that non-interference of psychological and physical explanation warrants our belief in the compatibility of mental and physical causes. But this is not the case. Non-interference of psychological and physical explanation only support a negative conclusion, that is, the conclusion that TQ need not exclude other types of causation. But it does not support the conclusion that TQ is compatible with other types of causation (His insistence on the inadequacy of TQ as model for psychological explanation only makes this point more evident). This conclusion requires an account of mental causation, but Burge fails to provide one.¹⁰

If this reasoning is correct, it shows that Burge cannot avoid the problems associated with the *No-overdetermination* principle without giving an account of mental causation and its relationship with physical causation. But devising such an account of causation for psychological explanation amounts to nothing less than collapsing back into MCD, the very thing that Burge tried to present as futile. To see that this is so, one needs only notice that a huge portion of the literature concerning MCD is explicitly devoted to developing accounts of causation which would satisfy these constraints (Crane 2006: 1124). Such are, for example, accounts of causation which appeal to counterfactual dependence (Loewer 2015), structural causes (Dretske 1988) or program explanation (Jackson and Petite 1990). Each of these aims to provide an account of causation which would make mental causation compatible with a physical model of causation.

¹⁰ This objection is similar to the worries Kim has with Burge's analysis: "The issue is how to make our metaphysics consistent with mental causation, and the choice that we need to make is between various metaphysical alternatives, not between some recondite metaphysical principle on the one hand and some cherished epistemological practice or principle on the other" (Kim 1998: 62).

The problem with Burge's attempt to present MCD as a misguided discussion can thus be formulated as a dilemma. If Burge does not provide an account of mental causation, he lacks a strong argument for the compatibility of mental and physical causation. On the other hand, if he tries to provide such an account, he ends up participating in MCD.

In addition, some portions of Burge's article can even be interpreted as occupying a position in MCD. Burge's central view, according to which the causal efficacy of properties depends on the way events are kind-individuated overwhelmingly reminds of the so-called *dual explanandum* versions of mental causation, according to which the mental and physical properties of an event are causally efficacious for different properties of the effect.¹¹ At one point, Burge says:

...we know that the two causal explanations are explaining the same physical effect as the outcome of two very different patterns of events. The explanations of these patterns answer two very different types of inquiry. (Burge 2007: 359)

or in the passage quoted in the previous section, when Burge says the following:

For the causal powers of a physical event that is mental might include possible effects that are specified in mentalistic explanation. (Burge 2007: 347)

If so, Burge's analysis inherits the problems associated with these kinds of strategies. Unsurprisingly, the main problem dual explanandum strategies face is the worry that they do not respect the *No-overdetermination* principle (Robb and Heil 2019), the very same principle, I argued, Burge cannot adequately overcome without giving an account of causation both appropriate for psychology and compatible with physical models of causation.

6. Conclusion

I will conclude by a brief methodological consideration. At several points in his article, Burge expresses his belief in some sort of meta-physical dependence between the mental and the physical:

There is certainly reason to believe that underlying our mental states and processes are physical, chemical, biological, and neural processes that proceed according to their own laws. Some such physical processes are probably necessary if intentional (or phenomenal) mental events are to be causes of behavior. (Burge 2007: 349)

On the other hand, however, his analysis of TIP and MCD supports a view of psychology and neuroscience as being importantly different scientific enterprises, whose taxonomies and explanations differ in important, even unbridgeable ways. The problem with such a view is that it leaves the relationship between psychology and neuroscience highly problematic. Since both of these are concerned with understanding cognition and behaviour, to insist that they can proceed on completely

¹¹ See Robb and Heil (2019) for an overview of such approaches.

separate courses would be to advocate something in the spirit of averroisian “double truth theory,” according to which it is possible for religion and philosophy to arrive at mutually contradicting but true knowledge. At one point, Burge seems to be fine with such a view:

Maybe science will never make use of anything more than limited correlations with the lower, more automatic parts of the cognitive system. Maybe identities or part–whole relations will never have systematic use. Maybe the traditional idea of a category difference will maintain a presence in scientific practice. (Burge 2007: 360)

Some philosophers argue for a different picture of the relationship between psychology and neuroscience. For example, Bechtel (2008: 71) argues that cognitive scientists use identities between mental and brain processes as heuristics which then serve to improve both the psychological and neuroscientific research. In a similar spirit, Polger and Shapiro (2016: 168–169), following Churchland (1986) see the relationship between the two as one of coevolution and interplay in which both behavioural experiments and neuroscientific manipulations converge in advancing our understanding of cognitive phenomena. If that is the case, as Burge himself seems to accept at one point (Burge 2007: 381), then the various problems which MCD identifies, such as the relationship between different causal models or the question of metaphysical dependency, far from being theoretically unmotivating, come out as important in understanding the relationship between psychology and neuroscience. However, what this relationship will turn out to be, I agree with Burge, is an open question.

References

- Bechtel, W. 2008. *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. London: Routledge.
- Burge, T. 2007. *Foundations of Mind: Philosophical Essays, Volume 2*. New York: Oxford University Press.
- Crane, T. 1995. “The Mental Causation Debate.” *Proceedings of the Aristotelian Society* 69: 211–36.
- , 2003. “Mental Causation.” In L. Nadel (ed.). *Encyclopedia of Cognitive Science*. London: Macmillan: 1120–1125.
- Churchland, P. 1986. *Neurophilosophy*. Cambridge: MIT Press.
- Dowe, P. 2000. *Physical Causation*. Cambridge: Cambridge University Press.
- Dretske, F. 1988. *Explaining Behavior: Reasons in a World of Causes*. Cambridge: MIT Press.
- Fodor, J. 1974. “Special Sciences: Or the Disunity of Science as a Working Hypothesis.” *Synthese* 28: 97–115.
- Jackson, F. and Petit, P. 1990. “Program Explanation: A General Perspective.” *Analysis* 50: 107–117.
- Kim, J. 1993. *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press.
- , 1998. *Mind in a Physical World*. Cambridge: MIT Press.

- Levine, J. 1983. "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly* 64: 354–361.
- Loewer, B. 2015. "Mental Causation: The free lunch." In T. E. Horgan, M. Sabates, and D. Sosa (eds.). *Qualia and mental causation in a physical world: Themes from the philosophy of Jaegwon Kim*. Cambridge: Cambridge University Press.
- Nagel, T. 1974. "What is it like to be a Bat?" *Philosophical Review* 83: 435–456.
- Papineau, D. 2000. "The Rise of Physicalism." In M. Stone and J. Wolff (eds.). *The Proper Ambition of Science*. New York: Routledge: 174–208.
- Polger, T. W. and Shapiro, L. A. 2016. *The Multiple Realization Book*. Oxford: Oxford University Press.
- Putnam, H. 1975. "The Nautre of Mental States." In H. Putnam (ed.). *Mind, Language and Reality: Philosophical papers*, vol 2. Cambridge: Cambridge University Press: 429–440.
- Robb, D. and Heil, J. 2019. "Mental Causation." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), forthcoming URL = <<https://plato.stanford.edu/archives/sum2019/entries/mental-causation/>>.
- Woodward, J. 2003. *Making Things Happen*. Oxford: Oxford University Press.