

## THE EXPERIENCE OF MENTAL CAUSATION

Jakob Hohwy  
*Department of Philosophy*  
*University of Aarhus, Denmark*

**ABSTRACT:** Most of us have a very firm belief in mental causation; that is, we firmly believe that our own distinctly mental properties are causally efficacious in the production of our behavior. This belief is dominating in contemporary philosophy of mind as a part of the causal explanatory exclusion problem for non-reductive materialists. I do not discuss the exclusion problem; rather, I assess the conception of mental causation that is presupposed in the current debate. I propose that in order to make sense of our firm belief in mental causation we need to operate with a broader conception of it than is normally seen, focusing on common-sense aspects concerning the timing, awareness, control, and tracking of mental causation. However, prominent studies in social psychology and cognitive neuroscience show that mental causation is not as self-evident, robust, and pervasive as our firm belief in it would suggest. There is therefore a tension between the common-sense, broad conception of mental causation and our empirical evidence for mental causation. A full defense of mental causation is not just a matter of securing causal efficacy but also of situating our notion of mental properties in relation to difficult issues concerning awareness, control, and judgment.

*Key words:* mental causation, conscious will, agency, social psychology, cognitive neuroscience

Many people, including most philosophers, have a very firm belief that there is mental causation, that is, that mental states such as beliefs, desires, intentions, and emotions are efficacious—*qua* instantiating *mental* properties—in the causing of some physical events such as bodily movements and actions in the wider sense. As Jaegwon Kim says, “some view [the existence of mental causation] as an ultimate, nonnegotiable commitment” (1998, p. 31). To top it up, Kim also quotes Fodor, who says that if there is no mental causation “then practically everything I believe about anything is false and it’s the end of the world” (Fodor, 1989, quoted in Kim, 1998, p. 32).

We thus have a very deep attachment to mental causation. Many people are also strongly attached to the belief that the physical domain is causally closed, with no causal over-determination, such that all physical events have purely physical

---

**AUTHOR’S NOTE:** This paper was presented at philosophy seminars at Aarhus University and La Trobe University. I thank the audiences on those occasions, and the editor of this journal and two referees for fruitful comments and suggestions. The research for this work was supported by a grant from the Danish Council for Humanities Research for the research project *Naturalised Mind—Cognisant Nature* (<http://www.namicon.au.dk>). Please address all correspondence to. Dr. Jakob Hohwy, Department of Philosophy, Aarhus University, DK-8000 Aarhus C, Denmark. Email: [filhohwy@hum.au.dk](mailto:filhohwy@hum.au.dk)

causes. But then there seems to be no room for causation of physical events by instantiations of mental properties *qua mental* properties. This is, in brief, the problem of causal explanatory exclusion for mental causation (Kim, 1998). Accordingly, an enormous number of people have tried to show how the belief in mental causation can be true even if the physical is causally closed. It is safe to say that no one has as yet succeeded in this (for core discussions see Heil & Mele, 1993; Kim, 1998).

This paper does not concern the very many accounts of how mental properties can avoid causal exclusion; rather, it concerns the conception of mental causation that people are so intent to defend against causal exclusion. This conception often seems rather poorly articulated, perhaps because it has such a strong intuitive pull. It is not always clear what the notion of mental causation that people are so firmly attached to has in common with the notion of mental causation that they try so hard to defend from causal exclusion.

My aim in this paper is twofold: (1) I propose a broader conception of mental causation than is usually seen in the debate; (2) However, under the broader conception our belief in mental causation comes under pressure from various prominent empirical studies in social psychology and cognitive neuroscience. A review of these studies provides evidence that mental causation is not as self-evident, robust, and pervasive as our firm attachment to it suggests.<sup>1</sup>

There is therefore a *tension* concerning the notion of mental causation; on the one hand it is analytic that we must operate with the broader conception of mental causation, but on the other hand there is empirical evidence that nothing actually fully satisfies this conception.

The upshot is, firstly, that it is not so clear just *how* it is “the end of the world” if there is a substantial threat to mental causation, and, secondly, that a solution to the problems surrounding mental causation cannot just consist in showing how mental properties can be causally efficacious but also in how mental causation relates properly to awareness, agency, and volition. Of course, the fact that mental causation is not as self-evident, robust, and pervasive as we believe does not show that there is no mental causation whatsoever, just as the fact that visual perception is not as reliable as many perhaps think (we can be taken in by visual illusions, etc.) would show that there is no visual perception whatsoever. However, it does show that there is reason to adjust our conception of mental causation, and it therefore leaves it an open question how the existence of mental causation is best defended.<sup>2</sup>

---

<sup>1</sup> Note that the Freudian notion of the subconscious could be used in a somewhat similar strategy about mental causation; however, that notion is much more narrow in scope than the modern focus on the “adaptive unconscious” (Wilson, 2002, Ch. 1), and would not target the notion of mental causation in the same way.

<sup>2</sup> The opposite view, that there is reason to adjust our empirical findings in the light of our conception of mental causation, is much more controversial and I will not consider it further here. Care is needed, however, because it is of course legitimate to *interpret* empirical findings, especially when they incorporate introspective reports, in the light of our conception of, for example, mental causation. That is, in order to understand what

The empirical studies I review could also be used to explore the philosophical notions of free will and of privileged self-knowledge. Indeed, some of them are explicitly presented as challenges to our everyday conceptions of free will and self. However, in my view it is advantageous to consider mental causation first. The study of free will must take compatibilism about free will and determinism as the default view,<sup>3</sup> and this is centrally a view about how some of our desires are causally efficacious in the production of our behavior. The study of privileged self-knowledge would benefit from making a clear distinction between self-knowledge of the contents of our mental states on the one hand, and of the causes of our behavior on the other hand, as well as from characterizing the scope of self-knowledge *vis à vis* the distinction between conscious and unconscious domains.

Of course, a staunch Cartesian, who believes in the unbridled transparency of the mental, would claim that notions such as the Freudian unconscious, and, by extension, scientific findings in psychology and neuroscience, should have no tendency to threaten our notion of the privileged domain of the mental. I think it is hard to give an *a priori* argument for such a position, and the contingent fact is that notions of the unconscious have thoroughly pervaded our conception of the mental, and of mental causation in particular.

It is an assumption in the argument to follow that we should examine the relation between phenomenology and behavioral sciences. A radical behaviorist might object that this is a spurious relation: behavioral science is *the method* for studying phenomenology, so there is not a third vantage point from which to investigate the “relation” between the two. I think my assumption is neutral on the issue of radical behaviorism. It is not so strong that it requires an independent third vantage point. The point is just that we have certain common-sense beliefs about mental causation and certain scientific beliefs about mental causation, and it is difficult to see how to reconcile the two sets of beliefs. The question of reconciliation can be addressed without assuming an independent viewpoint, e.g., by making explicit what common-sense beliefs are implicit in the scientific history, or what common-sense beliefs are implicitly ruled out by the scientific story. I think that this is not much different from the way our common-sense beliefs about heritability have been challenged and changed by scientific beliefs about the double helix.

### **The Conception of Mental Causation**

The belief in mental causation is often defended by referring to the obvious fact that we can very often predict what we will do and explain what we did in the mental state terms of common-sense folk-psychology, together with the fact that

---

subjects mean when they report their mental states it is useful to be guided by a sound grasp of their concepts for mental events.

<sup>3</sup> Though this is often ignored in favor of libertarian notions of free will, in which free action is seen as completely undetermined by the subject’s desires, etc.; libertarianism is an easy target because we only care about a free will that in some sense is responsive to at least some of our desires.

such predictions and explanations are naturally understood as citing causes of behavior—that is, the reasons for which we do what we do (Davidson, 1963; Kim, 1998). What makes this picture true is that there are, in fact, causally efficacious mental states. In support of these considerations of folk-psychological predictability is the firm conviction that it is *we* who are making many of the causal differences in the physical world—that we are not mere passive bystanders to the causal neurophysiological workings of our brains. It would be terrible for our self-understanding if we were sidelined in this way. If our mental properties were not causes, then there would be no agency or voluntary action (Kim, 1998, p. 31).

However, the belief in the causal nature of folk-psychological prediction and explanation, plus the belief that *we* make a difference in the physical world, is not sufficient to capture a conception of mental causation that is worthy of defense from causal exclusion. That is, there may be more to mental causation than just this rather narrow conception. Here is an illustration. In schizophrenia, some patients who have delusions of control (who say, e.g., “My fingers pick up the pen, but I don’t control them. What they do is nothing to do with me.” [Mellors, 1970]) have intentions to move, they move, know they move, and they move according to their intentions (see Frith, Blakemore, & Wolpert, 2000). So *they* are making a difference and they are folk-psychologically predictable and explainable, yet they attribute agency to someone else (see, e.g., Spence et. al., 1997). Now imagine we live in a world in which we all have delusions of control every time we intentionally move. In this world there would be much causation *qua* mental properties; *we* would in fact be making a difference, and there would be much actual folk-psychological predictability and explainability. It seems plausible, however, that there would not be much interest in defending mental causation against causal exclusion because each one of us would believe of ourselves that it is not *I* who is making the difference in the physical world: lots of people are making a difference, but it is never *me*! This would be very disconcerting, even though our own mental properties would be causally efficacious. Conceivably, many philosophers in this world would intend to *defend* the causal exclusion of the mental: perhaps the blind workings of neurophysiology are more acceptable than having one’s actions initiated by other people. This illustrates the point that in order to properly understand the presuppositions of the mental causation debate we need to consider a broader conception of mental causation.

Here are some of the commonly agreed upon aspects of mental causation:

**I. *Mental state predictability and explainability.*** Some physical events are predictable and explainable in the causal, mental terms of folk-psychology in the narrow sense of means–ends reasoning.

**II. *Reasons as causes.*** The mental cause of some physical events is the reason for which an agent acts.

**III. *Causal efficacy.*** Events can be causes in virtue of being instantiations of mental properties.

As illustrated by the story about the delusion of control world, this conception of mental causation is incomplete. We need to add some further aspects that concern agency and volition and our awareness thereof.

**IV. *Here-and-now causation.*** Our common-sense belief about action and our phenomenology of agency tell us that our beliefs and desires and their contents have an utterly immediate causal and explanatory relevance to action; they are current triggering causes (Horgan, 1991, pp. 88-89). Our experience of agency and folk-psychological practices would be radically different in a world without this here-and-now phenomenology.

**V. *Awareness of mental causation.*** In mental causation you are often aware of the mental property *M* that was causally efficacious in causing bodily behavior (i.e., you are often aware of the reason for which you acted). It is no good that there is mental causation if we are not often aware of it when it happens; in particular you would want to be aware of whether you were causally implicated in the action at all, and of the difference between acting for the reason given by the mental property *M*, rather than mental property *M\**, or whether the causally efficacious property was the non-mental property *P*. Part of the reason we are so impressed with the capabilities of folk-psychology is that we are aware of the reasons for which we act. Without such awareness we would have a very odd and indirect acquaintance with folk-psychology.

**VI. *Voluntary selection/endorsement and control of mental causes.*** You (mostly) voluntarily select, or at least endorse, the mental properties that are causally efficacious in the causing of bodily events, and you control the voluntary exertion of your body. It is no good having mental causation if you have no active hand in deciding or endorsing which mental properties are going to cause behavior and in controlling how they do it. We are not passive bystanders to neurophysiological happenings, but neither are we passive bystanders to mental happenings.

**VII. *Reliable tracking of mental causation.*** You generally reliably judge that there is mental causation when there is mental causation, both in our own case and in the case of others. If our folk-psychological practices are to make sense, then we must assume that our beliefs about mental causation reliably track mental causation. Though we are prone to be rather liberal in our judgments about other people, animals, and even inanimate things, it is no good if mental causation and the judgments about mental causation can come radically apart. Thus, for example, mental causation worth caring about cannot allow that folk-psychology enjoys predictive and explanatory success without being close to the truth about mental causings.

Now follow a number of comments on this account of the broad conception of mental causation.

I have separated out the last three aspects concerning awareness and volition in three parts (V–VII) because these three elements seem *prima facie* dissociable; there could be awareness without a feeling of agency, there would be a feeling of agency without reliable judging, and there could be reliable judging (e.g., via inference to the best explanation) without a feeling of agency or direct awareness of mental causation.

The beliefs set out in I–VII comprise a core part of our common-sense conception of mental causation. The claim is not that we are *always* aware and in control of mental causation, or that it is always here-and-now, or that we always track it perfectly. We know enough about different psychological notions and mechanisms (e.g., the Freudian subconscious and the workings of psychotropic substances) to make us accept less than total access to, and control of, mental causation; however, we are not prepared to countenance systematic and profound undermining of these aspects of mental causation. The claim is that these beliefs dominate our conception and experience of mental causation and that they are under-described in the literature. Take away any of these aspects and what is left is not really recognizable as the conception of mental causation that we so firmly believe in and that we are so intent on defending.<sup>4</sup>

How would I respond to someone who disagrees that we are committed to the broad conception of mental causation? I have not presented empirical evidence that we are so committed but have presented it as a piece of common-sense, armchair reasoning on our concept of mental causation. In this sense my claim is that the broad conception is analytic, a matter of conceptual analysis in a broad sense (see Jackson, 1998 for a full-scale defence of this notion of analysis). Disagreement would therefore concern conceptual differences. Some people, myself included, understand mental causation in terms of the broad conception. Other people may understand it in terms of another narrower or broader conception. It is then an empirical question as to who is in the majority and how precise my account of the broad conception is. My intuition is that most people would agree that the imagined worlds, in which one or more of IV–VII are false, are not worlds in which we have mental causation in the normal sense of that notion.

I said that the narrow conception associated with I–III to a large extent rests on the mental state terms and practice of means–ends reasoning of folk-psychology. The broad conception, in contrast, relies on the wider phenomenology of mental causation. This does not imply that there is some schism between the two conceptions in respect of the status of folk-psychology. I think that the broad conception’s dealings in notions such as voluntary control, timing, and awareness are just as much part of folk-psychology as the narrow conception’s dealings in notions of predictability and explainability. Part of my point is that this is easily overlooked in the debate about mental causation. Moreover, IV–VII is perhaps

---

<sup>4</sup> There might, of course, be further aspects, e.g., concerning the possibility of knowledge (e.g., Kim, 1998, p. 31) and phenomenal belief (e.g., Chalmers, 1996).

more closely related to our subjective experience but is nevertheless reflected in folk-psychology. I am not sure that I–III makes much sense without IV–VII and vice versa. Our practices of means–ends reasoning in mental state terms would not be the same if they did not rely to some extent on the notion of control, awareness, timing, and tracking. Our ideas of control, etc. would also have very little force if they did not relate to some extent to our practices of means–ends reasoning in mental state terms.

Lastly, I–III might, of course, be true without IV–VII being true; then there would be causation in virtue mental properties. So there is *a sense* in which there can be mental causation without aspects of awareness, agency, volition, and tracking. The point here is that this is not likely to be the kind of mental causation we care about.

### **Evidence for the Belief in Mental Causation**

Our belief in I–III will be taken for granted here. Belief in IV–VII is sustained by the wider phenomenology of mental causation. When we do act for reasons it seems to work immediately; we often experience awareness of our intentions and experience voluntary control in the selection or endorsement of intentions and reasons, and it seems that our judgments about mental causation mostly track the real thing. These experiences of mental causation also provide part of our evidence for our belief in mental causation under the broader conception above: we believe in mental causation because we have these experiences (compare: what would our belief in mental causation be like if we did not have this kind of evidence but *only* evidence concerning folk-psychological predictability and explainability?).

However, some prominent empirical studies in social psychology and cognitive neuroscience erode the status of this evidence. They do not show that there is no mental causation under *any* conception, but a detailed review leads to the conclusion that the conception of mental causation that we care about defending is not as self-evident, robust, and pervasive as we believe.

Before looking at these empirical studies that concern our evidence for the broad conception of mental causation, we should briefly consider the status of empirical, third-person research in this area. There are two potentially problematic notions, namely “mental” and “causation.”

Concerning “mental,” there is, beginning with behaviorism, widespread skepticism about the susceptibility of subjective (or phenomenal) phenomena to third-person science. The main problem is that there seems to be no independent way to ascertain the accuracy of introspective reports and thereby the experimenter’s interpretation of those reports. I shall simply assume that there are no *special* problems concerning the reliance of introspective reports in the studies I review below.<sup>5</sup>

---

<sup>5</sup> For a recent discussion of this complex and much-debated issue see the recent papers in *Trusting the Subject*, a two-volume special issue of *Journal of Consciousness Studies*, edited by Tony Jack and Andreas Roepstorff (2003, 2004). My own view, set out in Hohwy

Concerning “causation,” there is considerable difference between the relatively lax everyday judgments of causal relations and the definition of causation in experimental science in manipulationist terms of independent production. Of course, from a philosophical perspective there is even more debate about the correct analysis of the concept of causation. In most of the studies reviewed below, the issue is whether or not participants’ judgments of mental causation come apart from the facts of causation when the latter is manipulated by the experimenter, so there ought to be no general problem concerning causation in these studies.

A special problem may arise because in some cases the stimulus presented by the experimenter causes the participant to perform an action (e.g., the stimulus may be an instruction to raise an arm). From a radical behavioral point of view on such cases, what could make us distinguish causation by the stimulus from causation by the subject’s own mental states? Notice first that this problem does not arise easily in experiments, such as Wegner and Wheatley (1999), reviewed below, in which the stimulus concerns one thing (e.g., “manipulate a computer mouse”) and the action up for attribution concerns something else (e.g., whether the cursor on the screen stops in a certain place). In that case the stimulus causes one type of action and the attribution of mental causation concerns another action. The problem does, however, arise in other experiments (e.g., Libet et. al., 1983, also reviewed below) in which subjects are instructed to exercise their free will at will. A general solution to the problem depends on our ability to make sense of a distinction between two sorts of causal routes from stimulus to action: (a) the subject acts blindly, non-consciously on some stimulus, and (b) some stimulus causes the subject to form a desire that she then acts on. The distinction will attain some experimental credence to the extent we have evidence that (a) and (b) could be dissociated. However, it should be acknowledged that experimental dissociation will not get off the ground unless some trust is placed in subjects’ introspective reports, which takes us back to the problem of the use of such reports in experimental science. Still, there are clever ways to investigate mental causation in addition to simply asking subjects. Haggard, Clark, and Kalogeras (2002; reviewed below) contrasted voluntary action with movement induced by transcranial magnetic stimulation and discovered subtle temporal discrepancies between actual and experienced events, strongly indicating a difference between (a)- and (b)-type events. Blakemore, Oakley, and Frith (2003) discovered that hypnotized subjects experienced their agency differently in two conditions in which they themselves moved their arm as an effect of a stimulus. In one condition they were instructed to move their arm; they did, and they felt it was voluntary. In another condition they were given the hypnotic suggestion that their arm would be moved by a pulley system; they responded by moving the arm themselves, but they felt it was not voluntary. PET imaging revealed that the feeling of involuntariness is correlated

---

and Frith (2004) in volume II of that special issue, is that discovery of the correct interpretation of introspective reports goes hand in hand with successful explanation of the cognitive mechanisms underlying conscious states.



with increased activity in cerebellum and areas of the parietal cortex. Here, two different stimuli give rise to the same controlled movement but different subjective experiences, strongly suggesting that different causal routes from stimulus to action play a role in conscious mental causation.<sup>6</sup>

We now turn to the empirical studies that threaten our evidence for the broad conception of mental causation.

### *Apparent Mental Causation*

Wegner and Wheatley (1999; expanded in Wegner, 2002) developed a theory of apparent mental causation according to which “people experience conscious will when they interpret their thought as the cause of their action. This means that people experience conscious will quite independently of any actual causal connection between their thoughts and their actions” (Wegner, 2002, p. 64). Three factors are sufficient for producing the experience of conscious will. **Priority**: the thought should precede the action at a proper interval. The apparent mental cause event needs to occur just before the effect event to be perceived as the cause (notice that priority is consistent with the here-and-now character of mental causation mentioned in IV). **Consistency**: the thought should be compatible with the action. People generally expect some kind of similarity between cause and effect, though it is hard to specify a general account of what those similarities should be. **Exclusivity**: the thought should be the only apparent cause of the action. If there is reason to believe that other agents are in play, then the feeling of conscious will is diminished. Notice that the theory is very weakly specified; it does not mention intentions to do specific things, only thoughts about objects or events.

The theory predicts that subjects can experience conscious will even though there is no real mental causation. Wegner and Wheatley (1999) tested this prediction by making an experiment inspired by the Ouija board game. They instructed a participant and a confederate to randomly manipulate a computer mouse together for about 30 seconds at a time while listening to instructions over headsets and looking at the computer screen. The participant would be primed with words for objects, some of which were seen on the computer screen’s picture of a number of little objects. It was first established that priming the participant did not cause him or her to move the cursor any closer to the mentioned object than when the object was not mentioned. Next, the confederate would, on some trials, be instructed to make the cursor stop at a particular object at a specified time, and the participant would be primed with the word for this object either 1, 5, or 30 seconds before the stop, or 1 second after the stop. In this condition the participant’s thought about the object, induced by the priming, was not causally efficacious in making the cursor stop at that object. Nevertheless, the participants reported an increased experience of intention for stops when they had been primed 1 to 5 seconds earlier (mean percentage rated intentionality of about 60%). The

---

<sup>6</sup> I thank the editor and two referees for bringing up some of these problems.

experience of intention was rated as considerably lower when participants were primed 30 seconds earlier or 1 second after (mean percentage rated intentionality of about 45%).<sup>7</sup>

In this study people think they are performing an action that they are in fact not performing. It shows that just having a thought about an object at the right time is sufficient to make people experience that their thoughts are, to some extent, causally efficacious in the production of an event involving that object. Although the study does not show that there is no mental causation whatsoever (after all, the confederate's mental states seem causally efficacious in producing the stops), it does touch on various aspects of the conception of mental causation.

It is a commonplace idea that we are good, but far from infallible, at attributing intentional action to others. We sometimes accuse people of doing something intentionally when in fact it was unintentional happenstance, and we sometimes accuse people of something when it was a third person who was the culprit. When it comes to ourselves, however, we think that our judgments about our own intentional actions are pretty much correct (thus we tend to get insulted when we believe we did not do something on purpose and people accuse us of doing it intentionally, and similarly when we believe we did something intentionally and people suspect it was just luck). It is natural to think that this is because we experience awareness of intention in our own case but not in the case of other subjects, so we believe we are particularly good at tracking mental causation in our own case. The Wegner and Wheatley study shows that this may well be overconfident, though from this study it is hard to conclude anything about how often we get it wrong in our own case. Against VII, then, we must say that you are not as reliable a tracker of mental causation as you may think, not even in your own case.

The theory of apparent mental causation suggests a more general challenge to VII. If the theory is correct, then our judgments about mental causation are primarily tracking temporal contiguity of a thought about an object and a physical event involving that object. If this is what our judgments track, then it seems plausible that we at least cast the net too wide; presumably there is more temporal contiguity of this sort than there is mental causation.

If all that is required is something as vaguely specified as a mere thought about an object, then V is also challenged. The study suggests that belief in mental causation arises without awareness of the particular attitude toward the thought about the object. Was it an intention to move towards it, or away from it, or to tease the experimenter by going in circles around it, or was it just loose thought about what the experimenter could want to get out of this bizarre setup, or was it a thought associated with a desire to visually locate the object on the screen? (In fact, post-experimental interviews suggested the last one). This goes against the natural supposition that we believe in mental causation partly because we are

---

<sup>7</sup> Memory constraints could perhaps play a role in these results; however, given that the experience of intention also dropped sharply for priming just 1 second after the stop, memory does not seem to play a decisive role.

## EXPERIENCE OF MENTAL CAUSATION

aware of the content of our causally efficacious mental properties. This supposition would be part of V, but the evidence does not support it.

It also seems that we must question the evidence for our belief in VI concerning voluntary control. If our experience of intention is based on an inference involving temporal contiguity, then why should we think there is volitional control? After all, we do not think there is volitional control when a raindrop causes rings in a pond. It seems most plausible that the belief in voluntary control is also based on an inference concerning the likely interference of other agents, thus we would be most sure that events are under our voluntary control when there seems to be no one else around who could have produced it. This accords with the notion of exclusivity, mentioned above, and could explain why the mean percentage rated intentionality in the study did not go higher than about 60%; however, this kind of story is at odds with a natural understanding of VI. We do not think that an action was under voluntary control because we can see that there was no one there to interfere; rather, we believe in voluntary control because we think that we can consciously decide to perform certain actions. On this picture agency seems to be sidelined in a way inconsistent with VI.

Lastly, we should reconsider the evidence for the here-and-now phenomenology mentioned in IV. We might think that this aspect of mental causation arises because of something special in the experience of *mental* causation (e.g., our conscious awareness of it or our volitional involvement), but really it is a common feature of the experience of causation as such. In general, we experience causation partly because there is temporal contiguity between cause and effect, so the evidence for the here-and-now character is not particularly telling for the nature of mental causation *qua* mental.

The Wegner and Wheatley study is compelling but not entirely unproblematic. It consistently talks of conscious will or experience of intention, but nothing is done to distinguish these mental phenomena from a mere *inferred belief* in intention. It seems plausible that the participants came to the (60%) belief that they were the intentional agents behind the stop via an inference to the best explanation (drawing on premises such as that in the past thoughts about objects and actions involving those objects correlate with mental causation, etc.), but this may have happened without the phenomenology of an experience of intention or conscious will that we have in ordinary instances of action. Thus, what made participants rate intentionality at considerably less than 100% may not just be the modulation of the exclusivity parameter but the lack of some phenomenological component. Still, the study does show that priority, consistency, and exclusivity are more relevant for the feeling (or belief) of intentionality than is commonly thought.

### ***The Perceived Timing of Mental Causation***

Libet and colleagues (1983; Libet, 1985) used EEG to study the awareness of voluntary action in a series of still controversial studies (see *Consciousness and Cognition*, 11 [2002]). They found that the brain begins to prepare for voluntary

finger movement some 500 msec before movement, but that subjects become aware of “having the urge” to move only about 200 msec before movement, that is, well after movement preparation has begun. This has been taken by many, Libet included, to challenge free will: how can the will be free, if the brain is deciding when to do things before we have any inkling about what is going to happen? Some have pointed out that the studies are no threat to compatibilist conceptions of free will, which only require that we act according to (some of) our desires (Clark, 1999; Rosenthal, 2002). We may act according to some of our desires even if we are not aware of it as it happens, so it would be hard to conclude that these studies directly challenge the notion of voluntary control, as it occurs in V.

Libet also found that we are aware of having initiated the movement about 85 msec before the movement is actually initiated (before the finger begins to move). That is, we experience that the movement has begun well in advance of any actual movement (see also McCloskey et. al., 1983). This may be made possible by forward models of the movement that allows the prediction of the upcoming movement (Frith et. al., 2000). This is relevant for the perceived timing of mental causation. Firstly, it is interesting to note that the here-and-now phenomenology of mental causation mentioned in IV is partly based on an ability to predict events, that is, on “upcoming-events-happening-near-you” (cf. Spence et. al., 2002, p. 217). Secondly, the result is that the awareness of wanting to move and the awareness of having initiated the movement are much closer in time, about 100 msec, than the known neural events would predict. It may be that we have an active role in this temporal shift. Perhaps it would be harder for us to track mental causation if there was more than half a second between awareness of wanting to move and awareness of the movement happening—certainly this would be very odd (this explanation is suggested in Frith, 2002). If this is the case then we actively boost the experience of priority (one aspect of Wegner and Wheatley’s theory of apparent mental causation) and with it the here-and-now character of mental causation.

Temporal manipulation is found elsewhere. When a subject’s voluntary action is followed by some consequence (e.g., a tone) there is an attraction effect (called intentional binding) in the perceived times of the action and the consequence. That is, the action is perceived to be closer in time to the effect than it really is. The effect is reversed when the action is involuntary (such as muscle twitches induced by transcranial magnetic stimulation [TMS]) and is modulated by temporal contiguity and predictability (Haggard et. al., 2002). This again suggests that priority and the here-and-now character are boosted by us. In fact, to some extent the here-and-now experience arises because we believe that there is mental causation. It is interesting to note that when there is no mental causation (the TMS condition) we expand the time between movement and effect. Again, our experience of mental causation may causally influence our experience of priority.

We also sometimes fail to have awareness of why we respond as we do until after we have responded. Castiello, Paulignan, and Jeannerod (1991) had subjects grasp an object that was suddenly displaced. They corrected their movement accordingly but reported awareness of the displacement some 300 msec after onset

of the movement correction. They were aware of why they corrected their movement (they moved because the target was displaced), but only after they had moved. This case should make us beware of overstating the notion of conscious control (VI). The subjects felt in control but they did not consciously select the specific movement in response to the displacement, though they did some time earlier form the intention to grasp the object when it was displaced. Here there is at best endorsement of the mental cause, and the here-and-now experience would be based on a much earlier intention or a somewhat later awareness rather than on experience of the actual temporal sequence of events.

These studies impact IV. On the one hand they suggest that the here-and-now experience is significantly boosted by our belief in mental causation, consistent with the idea that mental causation is based on merely having thoughts about the effect event. On the other hand, even when subjects can predict the effect, that is, have a thought that a tone follows a muscle twitch (in the TMS condition in the Haggard et. al., 2002 study), they fail to believe that they were the agents and the perceived interval between action and effect is expanded. It seems, therefore, that there must be some underlying cognitive mechanism that labels our intentional actions and enables the contraction of events that gives rise to the experience of close temporal contiguity of mental cause and physical effect.

### ***Facilitated Communication***

In the section on apparent mental causation we considered cases in which subjects believed that they caused events that they did not cause. Wegner and his colleagues have also studied cases in which subjects believe that they did not cause events that they did in fact cause (for a review see Wegner, 2002, Ch. 6). For a short time it was believed that people who had grave disorders of communication (e.g., the low-functioning end of the autism spectrum) were able to suddenly communicate proficiently via a facilitator who held their hand while they typed on a keyboard. This belief was false: it was the facilitator who did the typing (in a decisive study the facilitator and the disabled person heard different questions and the typed answers were exclusively to the questions put to the facilitator; Wheeler et. al., 1993). The facilitators generally had the strong belief that they were not voluntarily typing, which goes against V–VII: the facilitators were not aware of the mental properties that were causally active in their actions, their actions were not consciously controlled, and they were poor trackers of mental causation.

Wegner and Fuller (reported in Wegner, 2002) had participants, acting as “facilitators,” put a finger on a “yes” key and a finger on a “no” key, and confederates, acting as “communicators,” rest their finger on top of the participants’ fingers. Both groups were issued headsets and a series of questions were posed to the facilitators, who were led to believe that the communicators were asked the same questions, whereas in reality no questions were posed to the communicators. The facilitators were instructed to read the communicators’ “unconscious finger movements” and press the key the communicators would have pressed. The facilitators answered correctly 87% of the time overall (in one

condition they did not even touch the communicator, but simply looked at him or her at a distance; in another they were told to respond randomly, yet they got 82% correct), and they rated the communicators' influence on their responses at 37%. This is surprising since the communicator could not have "communicated" anything else than random movement.

This study indicates the same pattern as seen in facilitated communication: subjects fail to be fully aware of the mental properties that are causally active in the production of the correct answers, they fail to reliably track the only mental causation that occurs (namely their own), and their choice of causally active mental properties is not fully under voluntary control since they manage to use prior knowledge to answer questions without wanting to. This further erodes the evidence for mental causation as captured by V–VII. Wegner suggests (2002, pp. 206-208) that it is the mere belief that other agents may be causally involved that make subjects project their agency to other agents. Thus, when subjects were primed with a skeptical belief about the validity of facilitated communication, their rating of the communicators' influence decreased somewhat and vice versa for priming with positive beliefs about the process.

Evidence for VI, concerning voluntary control, were further eroded in one condition in which facilitators were told to counteract the communicators' movement by applying an upward pressure and found that accurate answers rose to 94%. Wegner explains this in terms of "ironic processes": the effect in which people who are trying hard not to do something end up doing it, especially if they are under some kind of mental load (Wegner, 2002, pp. 204-206). The effect is hypothesized to arise when the intention not to do X nevertheless prompts a mental search for X-features that then can surface, especially under mental load. Ironic processes have the effect that we do the opposite of what we intend, so that when we consciously select one type of mental state ("press up, not down") as the cause of our actions, we end up doing the opposite.

It is difficult to say, on the basis of these studies, how widespread these effects and patterns of attribution are. Is it only in highly artificial laboratory settings that people project their mental causation, or is it common? Wegner suggests that it may be common, citing everyday instances that look like projecting mental causation (e.g., when people believe they have taught someone how to do something, but when this belief is based on projecting their own actions to a learner who in reality remains incompetent). The degree to which our evidence for mental causation is eroded would depend on how widespread these phenomena are.

Notice that the studies do show that even if mental causation is normally accompanied by a specific phenomenal sense of intention in addition to the experience of priority, consistency, and exclusivity, this sense can be easily trumped by changes in those three parameters. The mere belief that someone else could be responsible for the mental causation can make us believe that we ourselves are not responsible.

*Automatic Mental Causation*

We turn now to research on automaticity, in which subjects misidentify the reasons they acted as they did, or fail to identify any reasons at all, because those reasons were activated without involving consciousness. Automatic mental processes ensure that we can do things in response to environmental cues without the involvement of conscious attention. This is useful because conscious attention is slow and has a relatively limited capacity compared to automatic processes. Thus, the skilled driver on a familiar route may get home safely without paying much conscious attention to where she is going because nonconscious processes have taken over, leaving her free to think about other things. The range of automatic processes is, however, vastly larger than this well-worn example suggests.

Studies of the links between environment, perception, and behavior (reviewed in Bargh & Chartrand, 1999) suggest that they may often work entirely automatically. In studies of social perception it is found that social and racial stereotypes are automatically activated and lead to behavior influenced by that stereotype (e.g., people who are primed with the stereotype of the elderly [“Florida,” “sentimental,” “wrinkle”] walk slower and do not remember as well as controls). Likewise, people who are primed subliminally with photos of young, male African Americans react with more hostility to mild provocations compared with controls. Participants primed in this way also reacted with more hostility when playing games and rated their playing partners as being significantly more hostile than did non-primed controls—so they would be likely to misrepresent the reasons for their own hostility. In another paradigm, subjects automatically mimicked confederates when they rubbed their faces or shook their feet, and, conversely, evaluated the confederates who mimicked them as more likeable and easier to cooperate with than did controls who had not been mimicked. Importantly, in none of these studies were the participants aware that they were being primed, or of the influence of the priming.

Much research also focuses on the automatic activation of *goals* by environmental features. Such activation bypasses the ordinary notions of conscious, purposeful behavior associated with action theory and the broader conception of mental causation. Automaticity in this sense may be acquired intentionally, when one intends to acquire a skill (e.g., learning how to drive a car) or unintentionally, when the same choices are repeatedly made in the same type of situation. In both cases the frequent and consistent use of certain mental processes eventually leads to a removal of the conscious role in the process. If there is no conscious choosing of the goal then there should be no awareness of pursuing the goal either; subjects are not aware that their behavior is caused by their having those goals, and they cannot identify them in their judgments about mental causation. Bargh and Chartrand (1999) review studies that discern automatic activation of cognitive goals (e.g., the goal to remember information or to form an impression of someone; to restore one’s self-image [by, as it happens, denigrating others]); of automatic activation of behavioral goals (e.g., being motivated to

achieve and to have relative power in a situation); and also of automatic evaluation (e.g., picking negative or positive adjectives when primed with words such as “White,” “Asian,” “Clinton,” “tuna”), which turns out to be related to overall mood and avoidance behavior.

In a representative study (reported in Bargh & Chartrand, 1999, pp. 471-472) the behavioral goal of achieving well was primed in an allegedly unrelated first “word search” task in which words such as “strive” and “succeed” were presented to participants but not to controls. The participants duly outperformed controls in subsequent verbal tasks, but when interviewed they revealed no awareness of having their achievement goal activated.

Automatic activation of goals creates a state that is functionally very much like conscious activation of goals; for instance, participants primed with achievement persist despite obstacles or interruptions, and their subsequent mood is affected by success and failure in just the same way as for conscious activation (Bargh & Chartrand, 1999, p. 472).

The studies on automaticity affect the evidence for our belief in various aspects of mental causation. Contra V, for a very large number of our day-to-day intentional behaviors we are not in fact aware of the reasons for which they are done. Moreover, with respect to VII, when asked, we are wont to misidentify these reasons (thus we might say “yes, perhaps I behaved aggressively, but that’s because my playing partner was being even more aggressive,” whereas in reality the reason had to do with the activation of a racial stereotype). In these cases we may well be aware that we are engaging in intentional behavior such as playing a game, and we may sometimes correctly identify part of the reason (“the experimenter wants me to do this”), but a whole armory of mental states causally influence how this behavior proceeds without our awareness. We might say that in automatic behavior there may be reliable, albeit coarse, tracking of the occurrence of mental causation (sustaining a version of VII), but there is little awareness and tracking of the actual causally efficacious mental states.

Automatic activation fits poorly with the idea that there is voluntary control in the selection of causally efficacious mental properties (VI). If we are not aware of the activation of a goal then one can hardly say that acting on it is a voluntary choice. Of course, the goal may represent a longer-term desire that we do voluntarily endorse, but even this amount of voluntary control might be challenged by unintentionally acquired automaticity, where the process of automation is itself automatic. In those types of processes there may be consciousness in the early stages of pairing certain choices with certain situations, before the process becomes automatic, but the very acquisition of the underlying goal itself may not be intentional.

It is worth noticing that many automatically selected states occupy roughly the same functional roles as their consciously selected counterparts (e.g., achievers persist through obstacles) except that part of the functional role that involves reporting the reasons for one’s actions. This makes it plausible that much automatic mental causation of behavior remains folk-psychologically predictable and explainable. Much philosophical confidence in the reality of mental causation



## EXPERIENCE OF MENTAL CAUSATION

is based on the folk-psychological predictability and explainability of day-to-day behavior (witness the worn example of going to the refrigerator for a desired bottle of beer); these studies suggest that this confidence may rely on behaviors that are in reality automatic.

Automatic mental causation also seems relevant for the evidence for IV concerning the here-and-now character of mental causation. This evidence presumably comes from our experience of the constant conjunction of intentions and actions, but in automatic mental causation there is no such experience because we are unaware of the intentions (and sometimes also of aspects of the behavior that they give rise to). So for automatic mental causation there is no reason to insist on a here-and-now phenomenology, though the automatically activated states may, of course, be here-and-now causally efficacious.

Automaticity and its pervasiveness in our lives (Bargh & Chartrand, 1999, p. 463) pose no threat to the idea that mental properties can be causally efficacious. Automatically activated goals are mental representations (Bargh & Chartrand, 1999, p. 468), and when activated they produce their effect *qua* being mental properties (at least, there is no reason to deny that they are representations of how the agents desire the world to be, and as such they qualify as mental states). However, they do so without involving much of the aspects of mental causation described by V–VII; mental causation and our awareness and control of it drift apart.

As in the studies discussed above, it is hard to say how widespread automaticity is, though it seems to me that researchers are describing very familiar and frequent phenomena. It is therefore hard to assess how badly automatic mental causation erodes our evidence for the broad conception of mental causation—but there should not be any doubt that it does some damage. Consider the extreme, counterfactual version in which *all* mental causation is automatic. Here the causally efficacious mental properties have come completely apart from our awareness and voluntary control even though people will remain, to some extent, folk-psychologically predictable and explainable. This would, I believe, not be a world with a conception of mental causation that we can recognize at all, even though there would be much causation by mental properties.

I should also note the role of confabulation when behavior is driven unconsciously. Nisbett and Wilson (1977) presented subjects with four identical garments and asked them to choose the one of best quality. They preponderantly chose the right-most garment. The cause of this is that most people have a preference for things on the right, but the subjects reported that they felt the garment to be of a superior quality in some respect. Nisbett and Wilson (1977) cite a number of studies of this type. Wilson (2002, pp. 105-106) notes that the best explanation of this pattern of response is not that people never have access to their mental processes (as originally claimed in Nisbett & Wilson, 1977), but that when the causing is automatic or unconscious the subject cannot access the causes of behavior and therefore confabulates. It is a live empirical question to what extent people confabulate when there is a mix of conscious and unconscious causes in play, as well as *why* people confabulate.

One place to look for evidence is in split-brain experiments of patients in whom the corpus callosum normally joining the two hemispheres has been severed (for reviews and discussion see Gazzaniga, 1998, 2000). It appears that the left hemisphere confabulates when it does not have access to information given to the right hemisphere. For example, if the right hemisphere is given the command “Walk!” and the patient gets up and walks, the left hemisphere, which has language, will interpret the action as, for example, the manifestation of a desire for a can of coke. It is difficult to test, however, to what extent left-hemisphere confabulation is present in the healthy population (for a discussion see Gazzaniga, 1998).

The beliefs in V–VII are further eroded to the extent that we confabulate. In fact, the confabulations may be part of the cause of our having the strong belief in pervasive and robust mental causation in the first place. If, in cases of non-conscious behavior, we were generally disposed to say things like “I don’t know why I picked this garment” rather than “I picked it because it is much more elastic,” then it would be harder to make the case that any threat to mental causation approaches “the end of the world.”

### ***Fundamental Attribution Error***

We know that we are fairly liberal, and therefore fallible, in our mentalizing judgments about other beings; we over-use mental explanations (e.g., to certain animals and natural phenomena) and are not always right about other people’s motives. Still, we believe that we fairly often get it right when we try to discern mental causation in other people’s behavior, and certainly our own. This may be far from the truth. A series of studies have established that we are prone to commit what Ross (1977) calls “the fundamental attribution error”: we causally explain other people’s behavior in terms of their dispositional character traits (being generous, kind, hostile, honest, unfriendly, etc.) when in reality their behavior is caused by situational factors (being late, being made to believe that certain things are not dangerous, etc.).

Darley and Batson (1973) tested this in a study based on the Good Samaritan Parable (also described in Harman, 1999). In this parable a priest, a Levite, and a Samaritan pass a stranger in need, and only the Samaritan helps the man. The standard interpretation is that the Samaritan helps because of his religious character traits and the others failed to help because of their religious character traits—but perhaps other interpretations are more likely? A group of theology students with different religious backgrounds were asked to give a talk in another building. Some were told the talk should be on the Good Samaritan, others that it should be on other topics. Some were told that they were late and should hurry, some that they would be just on time, and some that they would be a little early. On their way they met a stranger in need. Sixty-three percent of subjects who were in no hurry stopped to help, 45% of those who would be just on time stopped, and only 10% of those who were in a great hurry stopped. It made no difference whether or not the students were assigned to talk on the Good Samaritan Parable,

nor did it matter what their religious outlook was. Thus, only the situational factor was relevant. Standard interpretations of the Good Samaritan Parable commit the fundamental attribution error of overlooking the situational factors, in this case overlooking how much of a hurry the various agents might be in (of course, these studies say nothing about judgments about people's character based on long-time observation in many different situations).

Again, we do not know how widespread the fundamental attribution error is. Could it really be true that our dispositional character never causes our behavior? Some think this is the case (see Harman, 1999 for discussion). It seems likely that it is fairly widespread and therefore that we often misidentify behavior as caused by conscious mental states when they are really caused by situational factors. Another interesting question is how often something like this happens in our own case (would we not say of ourselves that we stopped to help the stranger because we are nice people, rather than because we were not in a hurry?). There is evidence that we have a self-serving bias such that we attribute positive events to our character and negative events to situational factors (for asymmetries between first- and third-person attributions see Nisbett et. al., 1973; Watson, 1982). Ironically, when there *is* a very salient situational factor in play, such as a command from one's boss, then we tend to ignore the possibility that our dispositional character plays a causal role (e.g., that we would anyway desire to perform the ordered action; Wilson, 2002).

The causally efficacious situational factors would be mentally represented (e.g., the desire not to be late), so these studies pose no threat to the causal efficacy of mental properties. As we have seen above, however, they do erode our evidence in the beliefs set out in the broader conception of mental causation—our judgments about mental causation come apart from mental causation.

### **Consequences for the Broad Conception of Mental Causation**

I began by saying that some empirical studies may be relevant for our evidence for the broader conception of mental causation. After having reviewed some of these studies we must evaluate their overall impact. Firstly, they do not establish that philosophical epiphenomenalism is true (indeed, it is difficult to think of any empirical evidence that could accomplish this). All of the studies leave ample room for mental properties to be causally efficacious in various ways, and some even presuppose mental causation. This is important since some of these studies are presented as much more of a threat to mental causation. Wegner and Wheatley explicitly formulate their thesis as a threat to mental causation and they suggest that the actual causal route bypasses the mental altogether (Wegner & Wheatley, 1999, Figure 1, p. 483; Wegner, 2002, Ch. 9, pp. 68, 243-244; Wegner, 2003; see also Bargh & Ferguson, 2000; Frith, 2002). The Libet studies are often interpreted in similar strong ways; however, some of the studies are compatible with forms of psychological epiphenomenalism that do not hold that mental properties can make no causal difference in the physical world whatsoever (e.g., they can cause us to report them). Thus it may count as psychological

epiphenomenalism if conscious will is not causing behavior in the general sense, though it is causing behavior in the form of reports of experiencing conscious will. The studies do not permit us to conclude that universal psychological epiphenomenalism is true because none of them establish that conscious mental causation *never* occurs, though they do strongly suggest that it is very often disrupted or absent.

Underlying the discussion of epiphenomenalism is a more nuanced picture of the status of mental causation in the wake of these studies. I argued that the conception of mental causation that we care so deeply about must involve aspects that concern agency and volition (and our awareness of these) together with other phenomenological aspects and our ability to track mental causation. The studies show that mental causation under this conception is not as self-evident, robust, and pervasive as we may think and as some philosophical writings take for granted. Thus, our evidence for mental causation is eroded in several ways:

- We sometimes believe there is mental causation when there is none, often just on the basis of vague thoughts about the caused event.
- We sometimes fail to be aware of and identify mental causation correctly when it occurs, and even when we are on the right track we often only get part of the story, systematically misidentifying and ignoring large and important tracts of mental causes; instead, we confabulate about the real causes.
- Our evidence probably derives, in large part, from temporal contiguity of thought and physical event as well as from our confabulatory efforts, rather than just from some phenomenal sense of voluntary agency.
- Mental causation is often activated without our voluntary control/endorsement and awareness; intentions are occasionally acted on even though we have consciously decided not to act on them.
- It does not take much to disrupt the sense of agency and volition involved in mental causation; the mere belief that another agent is involved may remove the experience of intention. In fact, we may believe in voluntary control more because other agents are absent than because we have decided how to act.
- We artificially boost the here-and-now character of mental causation in various ways to make it seem more likely that we are the causal agents, even though we believe that we move before we actually initiate movement and even though we sometimes become aware of why we acted after we acted.

In general, then, there are many kinds of occasions in which the facts about causation by mental properties drift apart from our strongly avowed awareness, control, and tracking of mental causation.

This gives rise to the following tension. On the one hand, it seems analytical—a matter of armchair reflection—that we must operate with the broader conception of mental causation. If causation by mental properties is not closely

## EXPERIENCE OF MENTAL CAUSATION

tied to notions of awareness, control, and tracking, we can hardly recognize it as the concept we are trying to defend against causal exclusion (witness the imagined worlds in which one or other of the core beliefs about mental causation are false). On the other hand, empirical data do not bear out this broader analysis of mental causation; there is much less evidence than we think for such a close tie between causation by mental properties and awareness, control, and tracking. That is, if we think of the broad conception of mental causation as a cluster concept, or compound definite description, then the empirical evidence strongly suggests that nothing in fact *fully* satisfies the description.

In light of this tension we may have to downgrade the urgency of the problem of mental causation. Kim remarks that the seriousness of a philosophical problem depends in part on the depth of our attachment to the conflicting beliefs that make up the problem (1998, p. 30). We have a very deep attachment to mental causation, but we might legitimately ask whether, in the light of the empirical studies, we *ought* to have such a deep attachment to it. This downgrade would presumably only happen in step with a development in the way our common-sense beliefs reflect scientific findings.

On the other hand, Kim also remarks that the seriousness depends on how difficult it is to reconcile the conflicting beliefs. It will be a major achievement if a solution to the causal exclusion problem is found, but that might not give us peace of mind concerning mental causation because it is, as we have seen, caught up with messy notions of agency, volition, awareness, and judgment. It may be that a full defense and understanding of the causal efficacy of mental properties must await a resolution of important issues in the study of consciousness and thought. For example, mental properties must be causally efficacious whether they are consciously accessed or not; they must be able to fit into a temporal network in which goals can become non-conscious and then much later become automatically activated, and they must be the kinds of properties that allow us to perform manipulation of the perceived temporal relations between cause and effect.

The tension is also relevant to a certain deflationist argument in the debate about the causal exclusion problem. According to Burge (1993) there is no problem about causal exclusion because our reasons for believing in mental causation far outweigh our reasons for believing in the metaphysical considerations leading up to the problem. In a related vein, Baker (1993) argues that our widely accepted causal *explanations* in terms of mental states should be prioritized higher than the metaphysical notions of causation, and that then the exclusion problem simply melts away. Kim (1998, Ch. 3) argues against such deflationism, saying that serious philosophical problems hardly can be solved like that. The present discussion yields a more direct criticism. The deflationists *rest* their argument against the exclusion problem on the intuition that mental causation is self-evident, robust, and pervasive (they do not rest it merely on the intuition that mental properties cause us to report having them), so their argument is weakened to the extent that the empirical studies weaken this intuition.

We might ask what happens if the empirical findings are given the widest possible scope (cf., e.g., Bargh & Ferguson, 2000; Frith, 2002; Prinz, 1997;

Wegner, 2002, Ch. 9) such that mental causation is *always* apparent, non-voluntary, automatic, and temporally manipulated. Then I–III (the narrow conception) might be true (i.e., we would be folk-psychologically predictable and explainable and be making a difference because it would still be *our* mental states that are causes), the status of IV (the here-and-now aspect) would be somewhat vague, and V–VII (awareness, control, and tracking) would be false, although there would be cognitive mechanisms ensuring that it *seems* to us that IV–VII are true. One possibility in the conceptual landscape would then be the position according to which mental causation is not what most people would think (i.e., the broad conception). Rather, mental causation would just be constituted by the causal efficacy of causally determined conscious or nonconscious mental states in a way that allows folk-psychological predictability and explainability. The truth of the matter would not underpin our conception of mental causation, but we would have all the mental causation we care about because even though the phenomenology and the facts have come apart, this does not influence our folk-practices of prediction, attribution, and action.

Going for this position would not be the conservative option, and it depends on whether or not the studies can be generalized. Meanwhile, we can conclude conservatively that despite the very firm belief in mental causation, it is unclear how and if the day-to-day occurrences that are often cited as the obvious evidence for mental causation translate into the notion of mental causation that we are trying to defend. Put in Fodor's, we must ask: how near is “the end of the world”?

## References

- Baker, L. R. (1993). Metaphysics and mental causation. In J. Heil & A. Mele (Eds.), *Mental causation* (pp. 75-96). Oxford: Clarendon.
- Bargh, J., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*, 462-479.
- Bargh, J., & Ferguson, M. J. (2000). Beyond behaviourism: On the automaticity of higher mental processes. *Psychological Bulletin*, *126*, 925-945.
- Blakemore, S.-J., Oakley, D. A., & Frith, C. D. (2003). Delusions of alien control in the normal brain. *Neuropsychologia*, *41*, 1058-1067.
- Burge, T. (1993). Mind–body causation and explanatory practice. In J. Heil & A. Mele (Eds.), *Mental causation* (pp. 97-120). Oxford: Clarendon.
- Castiello, U., Paulignan, Y., & Jeannerod, M. (1991). Temporal dissociation of motor responses and subjective awareness. *Brain*, *114*, 2639-2655.
- Chalmers, D. (1996). *The conscious mind*. Cambridge, MA: Oxford University Press.
- Clark, T. W. (1999). Fear of mechanism: A compatibilist critique of “The Volitional Brain.” In B. Libet, A. Freeman, & K. Sutherland (Eds.), *The volitional brain: Towards a neuroscience of free will* (pp. 279-293). London: Imprint Academic.
- Darley, J. M., & Batson, C. D. (1973). “From Jerusalem to Jericho”: A study of situational and dispositional variables in helping behaviour. *Journal of Personality and Social Psychology*, *27*, 100-108.
- Davidson, D. (1963). Actions, reasons and causes. *Journal of Philosophy*, *60*, 685-700.
- Fodor, J. (1989). Making mind matter more. *Philosophical Topics*, *17*, 59-79.

## EXPERIENCE OF MENTAL CAUSATION

- Frith, C. D. (2002). Attention to action and awareness of other minds. *Consciousness and Cognition*, *11*, 481-487.
- Frith, C., Blakemore, S.-J., & Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London*, *355*, 1771-1788.
- Gazzaniga, M. S. (1998). *The mind's past*. Berkeley and Los Angeles, CA: University of California Press.
- Gazzaniga, M. S. (2000). Right hemisphere language following brain bisection: A 20-year perspective. In Gazzaniga, M. S. (Ed.), *Cognitive neuroscience: A reader* (pp. 411-430). Malden, MA: Blackwell.
- Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, *5*(4), 382-385.
- Harman, G. (1999). Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error. *Proceedings of the Aristotelian Society 1998-99*, *99*, 315-331.
- Heil, J., & Mele, A. (1993). *Mental causation*. Oxford: Clarendon.
- Hohwy, J., & Frith, C. (2004). Can neuroscience explain consciousness? *Journal of Consciousness Studies*, *11*(7-8), 180-198.
- Horgan, T. (1991). Actions, reasons, and the explanatory role of content. In B. McLaughlin (Ed.), *Dretske and his critics* (pp. 73-101). Cambridge, MA: Blackwell.
- Jack, T., & Roepstorff, A. (Eds.) (2003). Trusting the subject, Vol. I. *Journal of Consciousness Studies*, *10*, 9-10.
- Jack, T., Roepstorff, A. (Eds.) (2004). Trusting the subject, Vol. II. *Journal of Consciousness Studies*, *11*(7-8).
- Jackson, F. (1998). *From metaphysics to ethics*. Oxford: Oxford University Press.
- Kim, J. (1998). *Mind in a physical world*. Cambridge, MA: MIT Press.
- Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness potential): The unconscious initiation of a freely voluntary act. *Brain*, *106*, 623-642.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *The Behavioral and Brain Sciences*, *8*, 529-566.
- McCloskey, D. I., Colebatch, J. G., Potter, E. K., & Burke, D. (1983). Judgements about the onset of rapid voluntary movements in man. *Journal of Neurophysiology*, *49*, 851-863.
- Mellors, C. S. (1970). First-rank symptoms of schizophrenia. *British Journal of Psychiatry*, *117*, 15-23.
- Nisbett, R. E., Caputo, C., Legant, P., & Maracek, J. (1973). Behavior as seen by the actor and as seen by the observer. *Journal of Personality and Social Psychology*, *27*, 154-164.
- Nisbett, R., & Wilson, T. D. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review*, *84*, 231-259.
- Prinz, W. (1997). Explaining voluntary action: The role of mental content. In M. Carrier & P. K. Machamer (Eds.), *Mindscapes: Philosophy, science, and the mind* (pp. 153-175). Pittsburgh, PA: University of Pittsburgh Press.
- Rosenthal, D. (2002). The timing of conscious states. *Consciousness and Cognition*, *11*, 215-220.
- Ross, L. (1977). The intuitive psychologist and his shortcomings. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, Vol 10 (pp. 337-384). New York: Academic Press.

## HOHWY

- Spence, S. A., Brooks, D. J., Hirsch, S. R., Liddle, P. F., Meehan, J., & Grasby, P. M. (1997). A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control). *Brain, 120*, 1997-2011.
- Spence, S. A., Brooks, D. J., Hirsch, S. R., Liddle, P. F., Meehan, J. & Grasby, P. M. (2002). Alien motor phenomena: A window on to agency? *Cognitive Neuropsychiatry, 7*, 211-220.
- Watson, D. (1982). The actor and the observer: How are their perceptions of causality divergent? *Psychological Bulletin, 92*, 682-700.
- Wegner, D., & Wheatley, T. (1999). Apparent mental causation: sources of the experience of will. *American Psychologist, 54*, 480-492.
- Wegner, D. M. (2002). *The illusion of free will*. Cambridge, MA: MIT Press.
- Wegner, D. M. (2003). The mind's best trick: How we experience conscious will. *Trends in Cognitive Sciences, 7*, 65-69.
- Wheeler, D. L., Jacobson, J. W., Paglieri, R. A., & Schwartz, A. A. (1993). An experimental assessment of facilitated communication. *Mental Retardation, 31*, 49-60.
- Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious*. Cambridge, MA: The Belknap Press of Harvard University Press.