The Journal of Mind and Behavior
Spring 2007, Volume 28, Number 2
Pages 89–100
ISSN 0271–0137

Mental Action and Causalism

Jing Zhu

Sun Yat-sen University

This paper challenges the causal approach to understanding mental action by developing a pair of cases, both relevant to mental control. Central to the first case is the phenomenon of the ironic effects of mental control: our attempts at exercising control over our own minds can undermine the intended mental control itself. Central to the second case is the seemingly paradoxical notion of "passive mental action." These two cases indicate that the mental antecedents of the right kind specified by a causal theory of action are neither causally sufficient nor necessary to produce and control intentional mental action. This suggests that causalism may not be an adequate approach to understanding mental action.

Keywords: mental action, causalism, mental control

Some philosophers maintain that, just as there are bodily actions, there are mental acts or actions (Bishop, 1997; Geach, 1957; Mele, 1997a, 2000; Proust, 2001; Ruben, 1995). Actions are generally understood as things that people do, perform, or bring about, instead of things that people undergo or that merely happen to them. Many mental activities, such as making a decision, recalling a particular item from memory, doing a thought experiment, speaking silently to oneself, and solving a mathematical problem in one's head, are things that people actively do, rather than mere happenings. So it seems natural to consider these processes or events as mental actions.

There is a debate on whether a causal theory of action, or "causalism," the dominate position in contemporary philosophy of action since Donald Davidson published his influential paper "Reasons, Causes, and Actions" four decades ago

The author wishes to thank Andrei A. Buckareff for valuable comments and discussions. The Editor of JMB provided helpful feedback on the style of the original draft. Earlier versions of this paper were read at the third Columbia/NYU Graduate Conference in Philosophy and the 2003 Inland Northwest Philosophy Conference. I am grateful to Mark Couch, Noa Latham, Alfred Mele, and Scott Sehon for helpful comments. Requests for reprints should be sent to Jing Zhu, Ph.D., Institute of Logic and Cognition, Sun Yat-sen University, Guangzhou, China, 510275. Email: humanwill@yahoo.ca

(Davidson, 1963/1980), can accommodate mental action. Some philosophers are skeptical about whether the causal approach, which is primarily developed to account for overt, bodily actions, can be readily applied to the mental realm (see Cameron, 1991). Some philosophers challenge causalism using special kinds of mental action (e.g., "spontaneous" thoughts) as counterexamples (Ruben, 1995). On the contrary, causal theorists of action who take the challenges seriously argue that mental action poses no special threat to causalism, and can be readily accommodated by causal theories of action (Bishop, 1997; Mele, 1997a).

In this paper I challenge the causal approach to understanding mental action by developing a pair of cases, both relevant to mental control. Central to the first case is the phenomenon of the ironic effects of mental control: our attempts at exercising control over our own minds can undermine the intended mental control itself. Central to the second case is the seemingly paradoxical notion of "passive mental action." These two cases indicate that the mental antecedents of the right kind specified by a causal theory of action are neither causally sufficient nor necessary to produce and control intentional mental action. This suggests that causalism may not be an adequate approach to understanding mental action.

Causalism and Mental Action

According to causalism, what makes a bodily movement count as an action depends on the mental antecedents that have caused it. The central task of various causal theories of action is thus to specify what sort of causal antecedents is the appropriate kind: an agent's behavior can be counted as action if and only if it is caused in the right kind of mechanism by mental antecedents which constitute the agent's own reasons for the action (see Bishop, 1989; Brand, 1984; Enç, 2003; Goldman, 1970; Mele, 1992 for variants of an extended account of causalism). For example, according to Davidson's causal theory of action, an agent's having (1) a pro-attitude (typically a desire) toward an end, and (2) a belief that her action of a certain type can facilitate achieving the end, constitutes the reason for the agent to perform the action, *and* causes the corresponding movements of the agent's body (Davidson, 1963/1980).

Causal theorists of *mental action* claim that the causal approach can also be applied to understanding the nature of mental action. Alfred Mele (1997a) argues that various kinds of intentional mental actions, including remembering, imagining, deciding, and trying (making a mental effort), can all be given appropriate causal analyses, provided they exemplify a person's exercise of agency. He concludes that even "the weightiest challenges issued by mental action" can be surmounted, and mental action poses no "special, significant threat to a causal understanding of intentional action" (p. 247). Bishop (1997) believes that the best candidate for a naturalist theory of action is a causal

theory of action, and proposes that the causal approach can accommodate mental action by appealing to higher-order intentions, which allows adequate causal analyses of mental actions as well as rational, free actions to be derived from special kinds of mental action, such as decision-making. In what follows, I challenge causal theories of mental action by developing a pair of cases based on some striking features of mental control.

Ironic Effects of Mental Control

Most people can successfully exercise mental control in everyday life. We often effectively concentrate on working or studying, distract ourselves from unwanted thoughts, reasonably manage our emotions and moods, fall asleep as we want, hold back the impulse to be aggressive or socially unsuitable, and exert control over many other aspects of our mental life. In short, people seem quite adept at directing their thoughts, emotions, and motives according to their conscious intentions. However, we are also frequently thwarted in our attempts at exercising mental control. A large proportion of these failures are not like the errors that we inadvertently make, which could have been avoided or rectified had we devoted proper attention and efforts. Fairly often, exerted mental control can produce ironic effects, which are diametrically the opposite of what is desired. As Wegner observes:

> It sometimes seems that our desires to control our minds are met by an inordinate measure of failure. Whether we want to stop a worry, concentrate on a task, go to sleep, escape a bad mood, distract ourselves from pain, be humble, relax, avoid prejudice, or serve yet other mental goals, we find ourselves faltering again and again. Indeed, our attempts at mental control fall short so often that we may stop to wonder . . . whether there is some part of our minds, an imp of the perverse, that ironically strives to compel our errors. (1994, p. 34)

The ironies of mental control are that our very endeavors to exercise control over our own minds can yield precisely counterintentional results.

Ironic effects of mental control become especially dramatic when people are trying to *suppress* unwanted thoughts, moods, and motives (Wegner, 1989, 1992). In an experimental setting, people were asked to think aloud as they tried to suppress the thought of a target such as a white bear (Wegner, Schneider, Carter, and White, 1987). Subjects reported that they consciously tried to think of something else, in order to stop thinking about the target thought, and they could achieve considerable success. However, complete success was rare, because the thought to be suppressed returned again and again. Each time it returned, the subjects would try anew to distract themselves, and the cycle repeated itself. In such studies subjects typically report that they feel unusually sensitive to the unwanted thought throughout the period of suppres-

sion. In daily life, the experience of frustrated attempts of mental suppression seems abundant. The desire for eating or smoking can become even more vivid and salient when the person is aware of being in the process of dieting or quitting smoking. We may struggle to make particular thoughts go away in the midst of a sleepless night, only to have them return. When we try to forget a worrisome or unpleasant idea, the unwanted thought may even come back to haunt us more desperately.

Wegner argues that the mental processes that undermine the intentional control of mental states are "inherent in the very exercise of such control." According to him, "the ironies of mental life are not just happenstance examples of the frailty of human endeavors but rather are logically entailed by the nature of mental control" (1994, p. 34). To account for this phenomenon, Wegner proposes a model of mental control that contains two distinctive processes: (a) an operating process that promotes the intended change by searching for mental contents consistent with the intended state, and (b) a monitoring process that tests whether the operating process is required to work by searching for mental contents inconsistent with the intended state. The two processes support mental control by cooperative interaction. The operating process generates the desired change by "filling the mind with thoughts and sensations that are relevant to the desired state" (1994, p. 35). The monitoring process, which is sensitive to the failure of the operating process, detects whether the intended control has been achieved and so regulates whether or not the operating process will be triggered to promote the desired state. Whereas the operating process is effortful, attentive, and consciously guided, the monitoring process is usually unconscious, autonomous, and demands less mental effort.

According to Wegner's account, the operation of mental control usually requires certain amounts of psychological resources, including computing capacities, attention, efforts, and access to awareness. Normally, the operating process takes the lion's share of the psychological resources, which reflects the "current concern" or "identity" of the activity of mental control — the operation itself is what people are "doing" in exercising mental control. However, when the available psychological resources for mental control are substantively reduced for some reason (such as distraction, stress, time pressure, cognitive load, and so on), the operating process is subject to interference from the demands for mental resources by other mental processes and may even cease to function. The failure of the operating process triggers the monitoring process, which in turn will try to initiate the operating process. In this way, the monitoring process itself may surface to supersede the operating process in possessing the psychological resources and thereby overwhelm the intended control to yield the opposite of the mental state that is desired.

Mental Control and Causalism

It seems beyond dispute that mental controls are exercises of agency, by which an agent actively and intentionally does (or tries to do) things, rather than passively undergoes mental events that occur in his mind.[1] More specifically, we can define mental control as follows:

> An agent *S controls* her mind, which consists of a finite set of states and processes $\{m_1, m_2, \ldots, m_n\}$ in the normal range of S's mental capacities, if and only if the relation between S and her mind is such that S can effectively *drive* her mind into the state or process m_i $(1 \leq i \leq n)$ that S *wants* her mind to be in.[2]

It also seems uncontroversial that mental control is a component of mental action. Virtually every mental action contains a process of mental control, by which an agent purposefully changes or actively maintains the state of her mind.

Ironic effects appear strikingly peculiar to mental control. Physical efforts may fail as well, but their failures are most likely caused by physical restraints, rather than *undermined by the efforts themselves*.[3] In contrast, our thoughts seem remarkably capricious. The agony of mental control is that the whimsy always seems to haunt our attempts to direct our minds. A plausible general theory of mental action should make room to accommodate both our capacities to exercise mental control as well as the frequently experienced elusiveness and ironies.

Causal theorists of mental action claim that the causal approach is adequate for understanding mental action. According to causalism, an agent's mental event or process can be counted as a mental action if and only if it is caused by the right kind of mental antecedents which constitute the agent's own reasons for the action. For example, Alice wants to make a phone call to John. She believes that she remembers John's telephone number and that by performing the mental action of recalling, she is able to retrieve it from memory. Then she proceeds to retrieve the number. Her recalling the number is a mental action because it is caused by her desire to call John and her belief that a mental action of recalling enables her to retrieve John's number from her memory, which constitutes her reason for performing the mental action of recalling.

---

[1]See Dennett, 1984, ch. 3 and Bishop, 1989, pp. 22–25 for discussions of the role of agent-control in understanding human agency and action.

[2]This definition is based on one offered by Dennett for control *in general*:

> *A controls B* if and only if the relation between *A* and *B* is such that *A* can *drive B* into whichever of *B*'s normal range of states *A wants B* to be in. (Dennett, 1984, p. 52; see Zhu, 2005 for a discussion of behavior control and voluntary action.)

[3]Some behavior control, though, can also display ironic effects. The attempt to perform gracefully may diametrically result in awkward behavior. The effort to make an elegant public speech may make the speaker stammer.

Frankfurt (1978, p. 158) poses a classic objection to causalism:

> It is integral to the causal approach to regard actions and mere happenings as being differentiated by nothing that exists or that is going on at the time those events occur, but by something quite extrinsic to them — a difference at an earlier time among another set of events entirely.

Frankfurt argues that a more promising way of thinking about action is to consider the *concurrent* relation between an agent and her bodily movements. He suggests that an action is a bodily movement whose course is under the agent's guidance and control.

Since Frankfurt's article was published, however, causal theories of action have undergone notable developments. Causal theorists of action have largely discarded the ballistic conception of mental causation, according to which the causal contribution of the suitable mental antecedents does not extend beyond *triggering* an intentional action. Causal antecedents of action, either as reasons (e.g., appropriate combinations of beliefs and desires) or intentions, not only play a role in producing voluntary actions, but also have guiding, sustaining and controlling functions (see Adams and Mele, 1989; Audi, 1986; Bishop, 1989; Brand, 1984; Mele, 1992; Mele and Moser, 1994; Thalberg, 1984).

It seems straightforward for causalism to account for normal mental control. Not only do mental antecedents of the right kind (e.g., intentions) causally initiate the processing of mental control, they also causally guide and sustain the control process throughout. The causing mental antecedents play a central role in both the production and control of corresponding mental actions. Consider Susan, who wants to be happy. Her desire or intention to be happy kicks off the operating process to search for some joyful mental items (e.g., a joke or a cheerful song) that can make her feel happy. The monitoring process, which is sensitive to mental contents inconsistent with the intended state, detects whether she is already happy. If the monitoring process finds that she is still feeling sad, it will trigger the operating process to pursue a further search till she really feels happy or finally gives up the effort. Thus exercises of mental control, including mental efforts, can be given causal analyses in terms of causal history of the right sort (see Mele, 1997a for a causalist account of mental efforts).

Ironic effects of mental control apparently do not qualify as mental actions. Indeed, they are precisely counterintentional results of exercising mental control. Unlike other sorts of error or failure typically occurring in one's exercise of control, such as slips, side effects, or breakdowns due to external constraints, ironic effects inherently result from the exercise of mental control itself. They are responsive or sensitive to the contents and efforts of mental control. From the point of view of causalism, when an agent is exercising mental control, suitable mental antecedents are all in place — which should causally initiate,

guide and sustain the corresponding control process. But how is it possible for an agent not only to fail to generate the intended mental states or processes, but rather causally bring about the precisely undesired, counterintentional effects? The case of ironic effects of mental control appears to be a challenge to causalism.

Causal theorists of mental action may respond by pointing out that, in addition to mental antecedents of the right kind being in place, mental actions must be caused *in the right kind of way*. Normal exercises of mental control generate intended mental actions because they follow the right sorts of causal chains, whereas ironic effects of mental control are usually caused by mental antecedents of the right kind in *aberrant, abnormal* ways. This treatment, however, reminds us of the well-known problem of causal deviance, which generates certain types of counterexamples for causal theories of action.

A much-discussed example of deviant causation of behavior is Davidson's (1973/1980) unnerved climber. A mountain climber, who is holding up another climber, hopes to free himself from the danger he is in. He knows he can secure himself by loosening his grip. The desire and the belief unnerve him, and as a result he is caused to loosen his hold. The climber's mental states figure causally in the explanation of his behavior, but his unnerved behavior of loosening the grip hardly qualifies as an action.

A standard response of a causal theorist of action to the problem of causal deviance is to appeal to the idea that intentional actions must be caused by suitable mental antecedents *in the characteristic or right way* (Goldman, 1970, pp. 59–63; cf. Stout, 2002). However, there is by no means an easy way to specify what the characteristic mode of causation is by which suitable mental antecedents normally cause intentional action. After a thorough examination of various attempts to exclude cases of causal deviance and arguing that they are all unsuccessful, Bishop (1989, chs. 4 and 5; see also Peacocke, 1979, ch. 2) develops a sophisticated account to defend causalism from the threat of causal deviance. The basic idea is the sensitivity condition: non-deviantly caused behavior, which qualifies as action according to causalism, must show "a certain responsiveness or sensitivity to the content of the intention that causes it" (Bishop, 1989, p. 148). That is, the properties of the potential action must vary in appropriate ways as the intention varies. For example, Davidson's climber can intentionally drop the rope *abruptly* or *slowly*. These non-deviantly caused behaviors *are* actions because they are responsive to the content of the intention, whereas the climber's unnerved behavior is not. Bishop claims that his account can successfully exclude all cases of causal deviance (p. 172).

But the sensitivity condition is of no help for causalism to exclude ironic effects as mental actions, which are precisely responsive to contents and efforts of mental control. The lesson that I want to draw for causal theories of mental action is that it will be a no less challenging and difficult task to specify the

condition under which suitable mental antecedents *characteristically* trigger and sustain mental control rather than produce ironic effects. So the case of ironic effects of mental control indicates that the mental antecedents of the right kind are *causally insufficient* for the production and control of intentional mental action.

Passive Mental Action

In this section, I will show that not only the mental antecedents of the proper kind are not causally sufficient for a mental event or process to be counted as mental action, they are neither a necessary condition. Consider the following scenario:

> While Al is grading the essay on the trolley problem, a pleasant memory of a trolley ride in New Orleans comes to mind. The memorial thought distracts Al from his work and he tries to remember why he was in New Orleans at that time. He knows that Connie was with him on the trolley and he recalls that she accompanied him to a convention in New Orleans . . . . Memories (including memorial "images") of restaurants and shops in New Orleans, of the smell of the flowers in the French Quarter, of horse-drawn cabs, of stained glass windows in old churches, keep coming to mind.[4]

This scenario suggests an example of passive mental action. The seemingly paradoxical phrase "passive action" was coined by Mele (1997b) to describe a kind of action, introduced by Frankfurt (1978). A bodily movement is a passive action in virtue of the following features: (1) the movement is not initiated by the agent intentionally; (2) it is well under the agent's guidance and control, specifically, the agent is in the position to monitor the ongoing process, *and* is prepared to effectively intervene if necessary; and (3) the agent never *actually* intervenes in the course of bodily movement, because the proceeding of the movement is fully satisfactory. Frankfurt appeals to passive action in an attempt to undermine causal theories of action. Mele (1997b) argues that passive action does not constitute a special problem for a relatively standard causal theory of action.[5]

Is passive *mental* action possible? Consider Al, who intended to grade an essay on the trolley problem. He had no (prior) desire or intention to be engaged in the mental activity described in the above scenario. But when the memorial episode about his trolley ride in New Orleans begins to flood his mind, he takes it as a timely entertaining break from the tedious grading job

[4]This scenario is a selective amalgamation of two distinctive ones offered by Mele (1997a, pp. 237–238).

[5]Elsewhere I have argued that Mele's defense is unsuccessful and the case of passive action poses a real challenge to causalism (Zhu, 2004).

and decides to let it continue. He knows that he can stop the pleasant reconstructive mental activity and resume grading the essay. But at this moment he will rather enjoy the pleasure and put aside the grading work for a while.

It seems plausible to suppose that Al has the feeling of agency with respect to his mental activities. Even though he is not actively involved in the generation of the recollection of his vacation with Connie in New Orleans, he is apparently not a "helpless spectator in a bizarre mental theater" (cf. Mele, 1997a, p. 238). Once he feels that it is time to concentrate on working, he can readily stop his rumination. Of course, Al's desire to take a break and his belief that he can well manage to control the course of his memorial thought, substantively contribute to the generation of his feeling of agency. Suppose that, if he tried to turn his attention away from the thoughts of New Orleans and to direct it on the essay, but to no avail, he could hardly be qualified as an agent with respect to his thoughts. However, it is conceptually inadequate to think that the course of his spontaneous memorial episode must be *causally sustained* by his desire and belief, or intention. For one thing, what Al does is just to *monitor* the ongoing process and to be prepared to intervene once the progress does not proceed as desired. But it turns out that the causal mechanisms that stand ready to affect the course of thought never have occasion to do so. Monitoring is a unidirectional information process. It is by no means a sort of causal relation that can figure in the story that a causal theorist typically tells about the etiology of action. Furthermore, Al decides to allow the reconstructive mental activity to continue; if he decides to return to work, he can readily stop his daydreaming. But not stopping an ongoing process does not imply that one *causes* the continued processing. For example, Tom replaces Bill as a project manager. He finds that the plan worked out under Bill's direction is very good and the project's progress so far is entirely perfect. So he decides that he will not do anything to interfere with the execution of the project provided it runs well in accordance with the plan. Can we say that Tom's decision *causally sustains* the project's continued execution?

Of course, if Al had not made the decision to let the thoughts triggered by the association with a trolley ride continue, the mental wandering itself could hardly qualify as a mental action. And if he ceases to hold the intention to endorse the continued daydream, the ongoing mental activity can no longer be regarded as an intentional mental action. Nevertheless, Al's decision or intention to let his daydreaming continue is only *conceptually* or *analytically* necessary for his spontaneous memorial episode to be regarded as mental action, but not *causally* necessary. The case of passive mental action thus suggests that suitable mental antecedents specified by a causal theory of mental action are not *causally necessary* for the production and control of intentional mental action.

Conclusion

In this paper, I developed two cases against the causal understanding of mental action. The case of ironic effects of mental control indicates that causal analyses are insufficient to account for mental action. The case of passive mental action shows that suitable mental antecedents are neither causally necessary for the generation and execution of intentional mental action. These two cases jointly suggest that the causal approach, at least in its present forms, may not be adequate for understanding the nature of mental action.

Mental action is a prominent but poorly understood domain of human agency. The nature of mental action is relevant to our understanding of some other important aspects of our mental life, such as self-awareness, consciousness and the knowledge of our own mind (Peacocke, 2007; Soteriou, 2005). The causal approach aims to provide a straightforward and unifying perspective to understanding both bodily and mental action. But this paper casts some doubt on the outlook of the causal approach, and urges to take seriously some peculiar features of the mental realm, such as the ironic effects of mental control, in order to develop more adequate accounts of human agency.

References

Adams, F., and Mele, A.R. (1989). The role of intention in intentional action. *Canadian Journal of Philosophy, 19*, 511–532.

Audi, R. (1986). Acting for reason. *Philosophical Review, 95*, 511–546.

Bishop, J. (1989). *Natural agency: An essay on the causal theory of action*. Cambridge, England: Cambridge University Press.

Bishop, J. (1997). Naturalising mental action. In G. Holmstrom–Hintikka and R. Tuomela (Eds.), *Contemporary action theory* (Volume 1, pp. 251–266). Dordrecht: Kluwer.

Brand, M. (1984). *Intending and acting: Toward a naturalized action theory*. Cambridge, Massachusetts: MIT Press.

Cameron, J.R. (1991). Review of John Bishop, *Natural agency: An essay on the causal theory of action*. *Philosophical Books, 32*, 241–243.

Davidson, D. (1980). Actions, reasons, and causes. In *Essays on actions and events* (pp. 3–19). Oxford: Clarendon. (Originally published 1963)

Davidson, D. (1980). Freedom to act. In *Essays on actions and events* (pp. 63–81). Oxford: Clarendon. (Originally published 1973)

Dennett, D.C. (1984). *Elbow room: The varieties of free will worth wanting*. Cambridge, Massachusetts: MIT Press.

Enç, B. (2003). *How we act: Causes, reasons, and intentions*. Oxford: Oxford University Press.

Frankfurt, H.G. (1978). The problem of action. *American Philosophical Quarterly, 15*, 157–162.

Geach, P. (1957). *Mental acts: Their content and their objects*. London: Routledge.

Goldman, A. (1970). *A theory of human action*. Englewood Cliffs, New Jersey: Prentice–Hall.

Mele, A.R. (1992). *Springs of action: Understanding intentional behavior*. Oxford: Oxford University Press.

Mele, A.R. (1997a). Agency and mental action. *Philosophical Perspectives, 11*, 231–249.

Mele, A.R. (1997b). Passive action. In G. Holmstrom–Hintikka and R. Tuomela (Eds.), *Contemporary action theory* (Volume 1, pp. 135–143). Dordrecht: Kluwer.

Mele, A.R. (2000). Deciding to act. *Philosophical Studies, 100*, 81–108.

Mele, A.R., and Moser, P.K. (1994). Intentional action. *Noûs, 28*, 39–68.

Peacocke, C. (1979). *Holistic explanation: Action, space, and interpretation*. Oxford: Clarendon.
Peacocke, C. (2007). Mental action and self-awareness (I). In J. Cohen and B. McLaughlin (Eds.), *Contemporary debates in the philosophy of mind* (pp. 358–376). London: Blackwell.
Proust, J. (2001). A plea for mental acts. *Synthese, 129*, 105–128.
Ruben, D. (1995). Mental overpopulation and the problem of action. *Journal of Philosophical Research, 20*, 511–524.
Soteriou, M. (2005). Mental action and the epistemology of mind. *Noûs, 39*, 83–105.
Stout, R. (2002). The right structure for a causal theory of action. *Facta Philosophica, 4*, 11–24.
Thalberg, I. (1984). Do our intentions cause our intentional actions? *American Philosophical Quarterly, 21*, 249–260.
Wegner, D.M. (1989). *White bears and other unwanted thoughts: Suppression, obsession, and the psychology of mental control*. New York: The Guilford Press.
Wegner, D.M. (1992). You can't always think what you want: Problems in the suppression of unwanted thoughts. In M. Zanna (Ed.), *Advances in experimental social psychology* (Volume 25, pp. 193–225). San Diego, California: Academic Press.
Wegner, D.M. (1994). Ironic processes of mental control. *Psychological Review, 101*, 34–52.
Wegner, D.M., Schneider, D.J., Carter, S.I., and White, L. (1987). Paradoxical effects of thought suppression. *Journal of Personality and Social Psychology, 53*, 5–13.
Zhu, J. (2004). Passive action and causalism. *Philosophical Studies, 119*, 295–314.
Zhu, J. (2005). Voluntary action and behavior control. *Foreign Philosophy* (Issue 18, pp. 217–227). Beijing, China: Commercial Publishing House.