



# The Disappearing Agent and the Phenomenology of Agency

Jingbo Hu<sup>1</sup>

Received: 23 December 2023 / Accepted: 10 July 2024  
© The Author(s) 2024

## Abstract

The causal theory of action is thought to be plagued by the problem of the disappearing agent. However, philosophers have reached no consensus on the nature of this problem, let alone on whether it is solvable. In this article, I interpret the problem as a phenomenological challenge: the causal theory of action employs an event-causal framework, with which certain aspects of the phenomenology of agency seem incompatible. I examine two areas in which the phenomenology appears to speak against an event-causal framework: (1) the phenomenology of acting and (2) the phenomenology of making choices. Specifically, some philosophers have asserted that the phenomenology of acting involves content that action is caused by the agent themselves rather than by their mental states. Some philosophers have argued that the phenomenology of making choices involves content that the agent stands apart from their mental states in a way that allows them to reflect on, evaluate, or even manipulate their mental states. I argue that in neither case does the phenomenology pose a real challenge to the causal theory of action.

## 1 Two Readings of the Problem of the Disappearing Agent

Providing a proper account of human action is a difficult enterprise. The difficulty arises from a tension between two pictures of human action that we hold at the same time. On the common-sense picture, actions are very different from other natural phenomena—they are performed in an intentional and controlled way and are subject to rational and moral evaluations. On the scientific picture, by contrast, actions are not so different from other happenings—they fall within the natural causal order in

---

✉ Jingbo Hu  
jingbo.hu07@outlook.com

<sup>1</sup> School of Philosophy, Huazhong University of Science and Technology, Room 418, East Five Building, No. 1037 Luoyu Road, Hongshan District, Wuhan, Hubei Province 430074, China

which all causal interactions are event-causal (Bishop, 1989, 2010). Resolving this tension is a primary motivation for the causal theory of action (hereafter CTA)—a family of accounts that share the view that any given action is essentially caused in a specific way by certain motivational mental states or events. On the one hand, CTA elegantly distinguishes action from mere bodily movement (such as the knee-jerk reflex) by marking action, and only action, as being caused by specific motivational mental states in a specific way. On the other hand, CTA analyzes actions in terms of event-causation, a causal framework that is widely employed to capture other kinds of causal interactions in the world.

A simple version of CTA, often referred to as the Standard Causal Theory of Action (or standard CTA), contends that an action is the bodily movement of an agent resulting in a certain way from their motivational mental states, such as their beliefs, desires, and intentions.<sup>1</sup> Nevertheless, standard CTA is plagued by the problem of the disappearing agent because it contradicts a conviction entrenched in our conception of agency—the conviction that we actively engage in our actions. Call this conviction *Active Engagement*. Suppose I pick up a cup of coffee, ready to take a sip. According to standard CTA, this action amounts to my bodily movement (grasping the cup, raising it to my mouth), which is caused in a certain way by some of my motivational mental states (my desire to drink the coffee and my belief that I will drink the coffee when I have raised the cup to my mouth). In this story, I have dropped out of my own action because I have not done anything to make it happen.

However, there is no consensus on the nature of this problem, let alone whether it is solvable. To grasp the dispute, consider first the question of what an agent is supposed to do in the course of taking action. David Velleman, who has made significant contributions to the discussion of this problem, maintains that an agent has two roles to play in taking “full-blooded action”: (1) forming an intention to act via considering reasons for action and (2) executing the intention by moving the body (Velleman, 1992, p. 462).<sup>2</sup> According to Velleman, standard CTA fails to square with Active Engagement because it fails to accommodate these roles. A question remains as to exactly why standard CTA is incapable of capturing these roles. Different conceptions of the problem of the disappearing agent lead to different diagnoses.

Defenders of CTA are inclined to endorse a *weak reading* of the problem, contending that the agent disappears in the standard CTA story because the psychology of action presumed by the account is impoverished. Standard CTA says little about how an agent forms the mental states preceding their action (specifically, the process of deliberation) and why those mental states function on behalf of the agent (given that mental states can sometimes be alienated from the agent). This failure makes it difficult to account for how, during action, the agent makes decisions based on reasons. In addition, the account employs a sequential/ballistic picture of causation in

<sup>1</sup> It is doubtful that any philosophers actually endorse standard CTA, given the apparent problems it poses (such as the problem of causal deviance). However, it is common for some philosophers, particularly those who work outside the field of action theory, to use standard CTA as a working definition of action. To draw an analogy, few people believe that the JTB (justified true belief) analysis of knowledge is adequate. Nevertheless, many find the JTB analysis to be a helpful working definition of knowledge. Along these lines, it is safe to use standard CTA as a starting point for discussion.

<sup>2</sup> See also Steward (2012, 63) for similar remarks.

which the agent's mental states simply initiate the resultant bodily movement, neither sustaining nor guiding it.<sup>3</sup> This picture leaves no role for the agent in executing their intention.

Accordingly, CTA defenders believe that improved versions of CTA can make the agent reappear by providing more sophisticated psychology. One way to improve the theory is to incorporate a story about how an agent makes rational decisions within an event-causal framework.<sup>4</sup> Another way to improve the theory is to replace the sequential model of causation with more sophisticated models, whereby the agent's mental states not only trigger their bodily movements but also sustain and guide these movements. This could be achieved by positing a control system where the agent's mental states, bodily movements, and perceptions interact through feedback loops (e.g., Adams & Mele, 1989; Aguilar, 2012; Audi, 1986; Bishop, 1989; Thalberg, 1984), along with a hierarchical structure where different mental representations operate at different levels—personal and sub-personal (Pacherie, 2012).

By comparison, opponents of CTA do not think that such revisions genuinely solve the problem as they embrace a *strong reading* of the problem, according to which the event-causal framework shared by all versions of CTA cannot do justice to the role of the agent (e.g., Hornsby, 2004, 2008). On this reading, the only way to make the agent reappear is to ditch the event-causal framework entirely. Some philosophers propose as an alternative the agent-causal account of action, which holds that the essence of action is being caused by the agent themselves, not (or at least not *only*) by motivational mental states (O'Connor, 1995; Nida-Rümelin, 2007; Pereboom, 2015; Brent, 2017). Agent-causation is usually taken as a case of substance-causation, in which a substance figures as the cause in a causal relation. For example, the water dissolves the sugar; the flying stone breaks the window. Others propose a non-causal account of action, which takes “volition,” “trying,” or “will” as the locus of active control and maintains that this locus has no intrinsic causal structure and is therefore not subject to any causal analyses at all, let alone event-causal analyses (e.g., Ginet, 1990).

Although the strong reading makes the problem insurmountable for CTA, it is sometimes dismissed by CTA defenders as not being supported by substantial arguments. For example, Markus Schlosser once complained:

[Proponents of the problem of the disappearing agent] have not produced a single argument to support their case, and they have certainly not identified a philosophical problem. Their case is entirely based on intuition, and in some cases on mere metaphor and rhetoric. (Schlosser, 2010, 22)

<sup>3</sup> This critique is famously raised by Frankfurt (1978).

<sup>4</sup> This is basically how Velleman addresses the problem in his paper. Velleman thinks that the role of the agent can still be realized within the event-causal framework employed by CTA. He proposes adding to the original description of the action a psychological item, namely, the desire to act in accordance with reasons; this desire can play the role of the agent. See Velleman (1992). Similar strategies to address the problem can be found in Bratman (2001), Enç (2003), and Mele (2003). Shepherd (2022) even argues that a proper understanding of human psychology will dissolve the problem.

Indeed, the most common way to support the strong reading is by appealing to the anti-reductive intuition: the mere happening of things could never add up to an agent's active engagement. In an often-quoted passage, Melden expresses this intuition in the following way:

It is futile to attempt to explain conduct through the causal efficacy of desire—all that can explain is further happenings, not actions performed by agents... There is no place in this picture... either for rational appraisal or desires, or even for the conduct that was to have been explained by reference to them. (Melden, 1961, 128–9)<sup>5</sup>

But merely appealing to intuition is dialectically ineffective in supporting the strong reading; CTA defenders remain unpersuaded, as their position is motivated by the opposite intuition that Active Engagement is actualized within an event-causal process. To overcome the impasse, opponents of CTA wishing to make a case for the strong reading should do more than merely appeal to intuition.

The present article explores a potential way of substantiating the strong reading. Specifically, it approaches the problem of the disappearing agent as a phenomenological challenge, positing that the event-causal account of action is incongruous with certain aspects of our phenomenology of agency (Sect. 2). The investigation delves into two distinct types of phenomenology that ostensibly stand in opposition to the event-causal framework: the phenomenology of acting (Sect. 3) and the phenomenology of making choices (Sect. 4). Ultimately, however, it emerges that both these phenomenologies are, in fact, compatible with an event-causal account of action. In other words, CTA can withstand the scrutiny of the phenomenological reconstruction of the problem of the disappearing agent.<sup>6</sup>

Given the discussion that follows surrounding the dispute between event-causation and substance-causation, it is necessary to make a few observations regarding the ontological notions that are crucial to this discussion. Some may distinguish between *events* and *states*, with events—unlike states—being characterized by a change in properties. However, in this article, I will regard events and states as belonging to the same ontological category since both can function as causal relata within an event-causal framework. For example, the presence of oxygen (which is usually classified as a state) and the striking of a match (which is usually classified as an event) jointly cause the match to be lit. In certain contexts, I shall use the phrase “static events” to denote events (usually categorized as states) that appear to involve no changes in properties, and I shall use “dynamic events” to describe events that do involve changes in properties.<sup>7</sup> In treating events and states as belonging to the same onto-

<sup>5</sup> See also Nagel (1986, 110) and Taylor (1966, 109–111).

<sup>6</sup> A comparable recent approach to understanding the problem of the disappearing agent uses the analogy of the exclusion problem in mental causation. E.g., Franklin (2016, 2018); Himmelreich (forthcoming).

<sup>7</sup> Some authors may make an even more nuanced distinction between *states* and *events*. For instance, Steward (1997) argues that events and states have “different temporal shapes”. The motivation behind her distinction hinges substantially on linguistic considerations: different grammatical features are associated respectively with predications of “event” and “state”. However, the ontology of causation seems to be insensitive to these linguistic concerns.

logical type, I shall adopt Jaegwon Kim's (1966) conception of events as exemplifications of certain properties at specific temporal points.

## 2 The Pertinence of Phenomenology

Active Engagement is rooted in part in our first-person experience of agency: we take it that we actively engage in our actions because we *feel* that we actively make our decisions, execute our intentions, and move our bodies. However, when we are introspecting our first-person experience of agency, we find *prima facie* discrepancies between how our doings are felt and how they are described by an event-causal account. When we are executing an intention and moving our bodies, we feel as if our actions are caused by ourselves rather than by our motivational mental states. In addition, when we are making choices, we feel as if we are distinct from and manipulating our motivational mental states. Such discrepancies may easily elicit the intuition that the agent roles flagged by Velleman—forming and executing an intention—cannot be accommodated within an event-causal framework. Indeed, several proponents of the agent-causation account of action have remarked that their position is (in part) motivated by phenomenological concerns.<sup>8</sup>

The problem of the disappearing agent is, at first glance, a problem about ontology. How are concerns from phenomenology pertinent to a problem about ontology? To answer this question, let us first formulate the phenomenological challenge to CTA precisely:

**Phenomenological Challenge:** The phenomenology of agency sets up certain veridicality conditions that are not satisfied by CTA.

By “the phenomenology of agency” I mean the subjective experience the agent has that not only accompanies actions but also represents certain properties related to actions. Following Horgan et al. (2003), I take it that the phenomenology of agency sets up certain veridicality conditions that an action may or may not satisfy. I further take it that these veridicality conditions can be set up in two different ways. First, such a condition can be directly manifested by the phenomenal content. For example, if I see that an apple is red, my visual experience sets up the condition of the apple's being red. Second, a veridicality condition can be set up by the phenomenal content plus additional cognitive content. For instance, if I see that a man is old, my visual experience, together with certain background beliefs, sets up the condition of the man's having attained a certain age. Here, I grant that age is not a property that can be directly reflected in my experience. I “see” the man's age in the sense that I see the wrinkles on his face and the texture of his skin, from which, with certain background assumptions, I *infer* his age. Accordingly, the phenomenology of agency can challenge the event-causal account in two different ways. Either the phenomenal content involved directly manifests conditions that cannot be fulfilled by the event-causal

---

<sup>8</sup> E.g., Franklin (2016); O'Connor (1995); Pereboom (2015).

account, or the phenomenal content, taken together with certain reasonable assumptions, is incompatible with the event-causal account.

The phenomenological challenge outlined above, if established, constitutes a dilemma for CTA defenders: either the event-causal account is false, or our agential phenomenology is illusory. Still, some may find the second horn of the dilemma innocuous, suggesting that we should care about what action is but not what action is felt to be. In an unpublished manuscript (“Is the Phenomenology of Free Will Relevant to Its Metaphysics?”), Helen Beebe argues that it would not be especially puzzling or incoherent for someone who has the libertarian experience to endorse compatibilism. She writes:

We’re all used to things not being as they appear. Sticks look bent in water; redness really does look like it’s an intrinsic quality of objects; tables really don’t seem to be mostly empty space; and you can feel hungry and then, just as you’re reaching for a Tim Tam, realize that you can’t possibly be hungry because you only just had lunch. (Beebe, unpublished manuscript)

Although Beebe’s concern is about the pertinence of phenomenology to the free will debate, her reasoning readily applies to the current dilemma. If we can easily live with the fact that our perceptual experience is fallible, why can we not do the same with the phenomenology of agency?

To begin with, there is an important disanalogy here. The perceptual experience is generally reliable despite being fallible at times. By contrast, the phenomenology of agency, if shown to be illusory, would be illusory on an extremely large scale: we would be suffering from illusory experiences on every occasion of exercising our agency. Any account that fails to match our phenomenology of agency would become highly unattractive.<sup>9</sup> Relatedly, whether action should be captured by an event-causal framework is still a live and open question. The phenomenology of agency, if shown to be incompatible with CTA, would provide a *prima facie* reason against the event-causal account of action.<sup>10</sup> There may still be room to defend an event-causal account. But defending it would entail the burden of explaining why our phenomenology of agency is systematically non-veridical.

More importantly, the Phenomenological Challenge may undercut the primary motivation for CTA, namely to resolve the tension between the scientific and the common-sense pictures of action. The common-sense picture includes an important characteristic, namely Active Engagement, which is in part grounded in our phenomenology of agency. If it turns out that CTA is incompatible with the phenomenology of agency, that will mean that CTA cannot accommodate the characteristic of Active Engagement and thereby deviates from the common-sense picture of action. Put differently, acknowledging that CTA does not match our phenomenology is nearly the same as conceding that the problem of the disappearing agent (under the strong reading) is unsolvable. CTA would, at best, become a revisionary theory of agency. These

<sup>9</sup> For similar remarks, see Clarke (2019, 748–749); Nida-Rümelin (2007, 261).

<sup>10</sup> Cf. Wilson (2021, 277–278), who maintains that we should take the phenomenology of freedom at face value unless there are good reasons not to.

are the reasons why the phenomenology of agency is pertinent to the metaphysical concerns surrounding action and why defenders of CTA need to take the Phenomenological Challenge seriously.

### 3 Acting and the Experience of Self-As-Source

Let us focus on the second role of the agent cited by Velleman as problematic for CTA, namely the role of executing one's intention and moving one's body. That agents play this role is supported by a mundane experience we have whenever we do something by moving our bodies: we feel as if we ourselves, rather than our motivational mental states, are the cause of our actions. This phenomenology has been highlighted by Terry Horgan, who writes:

You experience your arm, hand, and fingers as being moved *by you yourself*—rather than experiencing their motion either as fortuitously moving just as you want them to move, or passively experiencing them as being caused by your own mental states... You experience your behavior as *caused by yourself*, rather than experiencing it as caused by *states* of yourself. (Horgan, 2011, p. 79, original emphasis)<sup>11</sup>

Horgan calls this phenomenological feature *self-as-source*. Rather than providing a straightforward articulation of this feature, he attempts to delineate it in a “negative/contrastive way”—by specifying what it is *not* like: it is *not* like feeling an occurrent wish for a specific bodily motion and then feeling the bodily motion take place passively, *nor* it is like feeling an occurrent wish for a specific bodily motion and then feeling a causal process consisting of the wish causing that bodily motion.

#### 3.1 Absence of Event-Causation vs. Presence of Non-Event-Causation

While people may be inclined to hold that self-as-source directly manifests a veridicality condition that can only be met by an agent-causal process and not by an event-causal one, Horgan himself is reluctant to draw such a conclusion. He holds that it would be better to construe the content of self-as-source as the *absence of event-causation* rather than as the *presence of non-event-causation*. With this distinction, Horgan maintains that self-as-source is still compatible with an event-causal account of action.<sup>12</sup>

But how exactly would such a distinction support the compatibility of our agentic phenomenology with an event-causal account of action? The best way to work it out is by invoking cases of reduction in science. Scientific progress has shown that many observable phenomena can be reduced to a more fundamental layer of reality. Take

---

<sup>11</sup> See also Horgan et al. (2003); Horgan (2007a, b). Other authors have made similar observations on the phenomenology of acting, e.g., O'Connor (1995), Nida-Rümelin (2007), Pereboom (2015), and Clarke (2019).

<sup>12</sup> Judisch (2010) has made a similar distinction to respond to this challenge.

as an example the identification of temperature with the movement of molecules. Our first-person experience of temperature is not presented as the movement of molecules—it is presented as the sensation of heat or cold—but nor is it presented as the negation of the movement of molecules. Because of this, the experience remains compatible with the reduction.

Along similar lines, if action is shown to be fundamentally underpinned by event-causal processes, then even if action is not experienced as event-caused, its phenomenology is still compatible with an event-causal account. This line of argument is adopted by Clarke (2019). Before illustrating Clarke's argument, let us first distinguish between the *ordinary sense* and the *fundamental sense* of agent-causation. An action is considered agent-caused in the ordinary sense if it is performed by the agent through their exercising certain capacities to act.<sup>13</sup> It is considered agent-caused in the fundamental sense only if the process of agent-causation is ontologically fundamental. Accordingly, the ordinary sense of agent-causation can be endorsed by everyone, including proponents of CTA, whereas the fundamental sense of agent-causation is genuinely incompatible with CTA.

Clarke's proposal, simply put, is that self-as-source supports agent-causation only in the ordinary sense; it remains epistemically possible that agent-causation in the ordinary sense is either reductively analyzable in terms of event-causal processes or non-reductively realized by them. Agent-causation in the ordinary sense is reducible to event-causation if event-causation is shown to be both ontologically fundamental and conceptually prior to agent-causation in action. It is non-reductively realized in event-causation if event-causation is shown to be ontologically fundamental but without conceptual priority.<sup>14</sup> As a result, even though self-as-source involves the content of agent-causation in the ordinary sense, it is compatible with the project of CTA, which tries to realize action in event-causal processes.

Clarke entertains some objections to his proposal. He admits that there may be concerns about either reducing agent-causation to, or realizing it in, event-causation. For instance, cases of causal deviance may indicate that actions cannot be reductively analyzed (a classic example of causal deviance in action can be found in Davidson, 1973), and some philosophers argue that all causations are fundamentally substance-causations (e.g., Lowe, 2008). Nevertheless, Clarke contends that such concerns come from arguments of metaphysics rather than phenomenology. In other words, even if these concerns turn out to be well-grounded, they would not make a case for the incompatibility between the phenomenology of agency and an event-causal account of action. Clarke also considers the suggestion that human phenomenology portrays our actions not only as if they are agent-caused but also as if they are agent-caused in an ontologically fundamental way. In response, Clarke asserts that a more plausible scenario is that our experience of agency remains neutral to these metaphysical inquiries. This is because our experience of agency is thought to be at least

<sup>13</sup> Clarke himself refers to this as “the restricted sense” of agent causation, in order to distinguish it from a more liberal sense of agent-causation that includes cases of non-action, e.g., a man, who accidentally stumbled and fell, caused the breaking of a vase (2019, 751).

<sup>14</sup> Clarke (2017) provides a scheme for either reducing agent-causation to, or realizing it in, event-causation. See also Ekstrom (2000, 114) and Lowe (2008, Ch. 6).



in part nonconceptual, and that the concept of agent-causation (in the fundamental sense) is of metaphysical sophistication.<sup>15</sup> Nevertheless, he acknowledges the possibility that some philosophers with relevant metaphysical expertise might perceive their own actions as being agent-caused in the fundamental sense because of the influence on their experience of theoretical convictions. He asserts that even if this is the case, the conflict between the experience and the event-causal account is attributable to extra-experiential representation. Besides, only a limited number of individuals, namely philosophers with special conceptual training, are likely to find their lived experience incongruous with the event-causal account. As a result, ordinary people's phenomenology of acting does not suffer from massive illusion.

### 3.2 From Phenomenal Difference to Phenomenal Negation

However, neither Horgan's nor Clarke's response serves as a final solution to the challenge posed by self-as-source. The incompatibility between self-as-source and an event-causal account can be established without the premise that self-as-source involves content representing non-event-causation. All that is required is to recognize that self-as-source radically *differs* from our experience of event-causation. To see this point, consider again typical cases of scientific reduction, such as identifying temperature with the movement of molecules. Our experience is neutral to these reductions, as the more fundamental layer of reality is not directly accessible to our experience, so our experience does not specifically conflict with it. However, the same may not be true of an attempt to realize agent-causation in event-causation, as these two types of causation may be associated with radically different experiences. The point is, if self-as-source is radically different from our experience of event-causation, then it would be suspicious to claim that agentive phenomenology is compatible with an event-causal account of action. In an exchange with Horgan, Martine Nida-Rümelin (2020) advances this point with a hypothetical scenario.

**Alexa's Transformation:** Alexa was once a normal agent. Like the rest of us, she experienced her actions as brought about by her self. However, she then underwent a strange transformation. Ever since, she has not experienced her actions in the usual way. Instead, whenever she acts, she feels that her actions are event-caused. For example, when she wishes to have a cup of coffee, she feels that her wish for a cup of coffee and her belief about how to get a cup of coffee jointly cause her relevant bodily movements. Alexa is shocked and finds this phenomenology alienating.

This case vividly demonstrates Horgan's contrastive/negative way of identifying the content of *self-as-source*. In the wake of the transformation, Alexa no longer has the

<sup>15</sup> Horgan (2007a, 2011) has also provided sophisticated arguments for the claim that we cannot ascertain the metaphysical conditions of the phenomenology of acting just by introspection (the conditions of concern to Horgan are broader, including agent-causation, causal closure, and determinism). Horgan's basic idea is that forming judgments as to the metaphysical conditions of the phenomenology requires conceptual competence, and people's conceptual capacity is limited. Thus, people cannot immediately exercise the relevant conceptual capacity to form a judgment while introspecting their phenomenology.

normal feeling of self-as-source in her actions. Instead, she feels that she is observing her mental states causing her body to move passively. Contrary to Horgan's suggestion, Nida-Rümelin argues that the best interpretation of the change in Alexa's experience is that the content of the experience of self-as-source includes non-event-causation. She introduces the following principle to facilitate the argument:

In general, if the change from experiencing something as having property F to experiencing that same thing as having property G necessarily involves an experience of change (the impression that the object itself has changed from having F to having G), then one may conclude that experiencing an object as having F includes experiencing it as not having G. (Horgan & Nida-Rümelin, 2020, p. 271)

This idea can be summarized as the principle that **Phenomenal Change Implies Phenomenal Negation (PCPN)**.

**PCPN:** If two experiences are different (with respect to the same *category* of properties), then the content of these two experiences is mutually exclusive (the content of one implies the negation of the content of another).

PCPN seems to be compelling. For example, the experiences of seeing something as square and seeing something as round are different with respect to object shape. According to PCPN, then, when we see something as round, we are also seeing something as *not* square, and vice versa. By the same token, if Alexa experiences a radical change when the experience of event-causation replaces self-as-source as her agentive phenomenology, then, based on PCPN, we can infer that the experience of self-as-source includes the negation of event-causation in its content.<sup>16</sup> As a result, self-as-source is compatible neither with an event-causal account of action nor with attempts to realize action within an event-causal process.

However, the argument from PCPN hinges on the idea that we *can* phenomenologically discriminate event-causation from non-event-causation. In other words, the argument presumes that we can experience a particular *type* of causation: that we can have the experience not only that our actions are caused but also that our actions are event-caused or non-event-caused. I find this presumption suspicious.

### 3.3 Discerning Event-Causation from Substance-Causation

To clarify, my quarrel is not with the distinction between event-causation and non-event-causation. Rather, it has to do with the nature of our access to that distinction.

<sup>16</sup> Nida-Rümelin points out a second challenge to CTA on the basis of **Alexa's Transformation**. Alexa finds the experience of her actions being caused by her mental states alienating. This is an ironic result for the event-causal account of action, for it is the alienating experience that is closer to the description of the event-causal account. Based on this result, Nida-Rümelin asserts that even if the experience of self-as-source is compatible with the event-causal account, it is compatible only in a superficial way. My discussion below focuses on the first challenge, although I take it that my strategy can respond to the second challenge as well.

Typically, we differentiate concepts in two ways. First, we can distinguish them conceptually by providing definitions, as we do with abstract concepts. For example, we distinguish “prime number” from “composite number” by defining the former as a number greater than one with only two factors (1 and itself) and the latter as a number with more than two factors. Second, we can differentiate between concepts phenomenologically by associating them with unique phenomenal properties. This is how we distinguish “pain” from “itch,” “sweet” from “sour,” “red” from “blue,” etc.

Event-causation and agent-causation are commonly differentiated in conceptual terms. Event-causation is defined in terms of the causal relata: both the cause and the effect are events. By comparison, agent-causation is taken to be a special form of substance-causation in which a substance, as opposed to an event, stands as the cause in a causal relation. It is important to note that the fact that two concepts are usually differentiated conceptually does not mean that they cannot also be differentiated phenomenologically. At times, the phenomenological and conceptual approaches to differentiating concepts overlap—we can distinguish a circle from a square both conceptually and phenomenologically. However, there are some concerns about the possibility of phenomenologically differentiating event-causation from substance-causation. (I shall move on shortly to the possibility of phenomenologically differentiating event-causation from agent-causation in particular.)

First, our visual experience seems to register no difference between event-causation and substance-causation. In everyday circumstances, event-causal and substance-causal statements are used interchangeably to describe the same phenomena. For example, we might state either that “the ball’s rapid flight caused the breaking of the window” (event-causation claim) or that “the ball broke the window” (substance-causation claim).<sup>17</sup> This interchangeability suggests that the identification of substance-causation or event-causation as such is underdetermined by our visual experience. This conclusion is not surprising, considering that events and substances are interconnected in our visual perception. When we observe an event, we generally see it as involving certain substances as participants; likewise, when we see a substance, we typically perceive it as participating in specific events (whether static or dynamic).<sup>18</sup>

Second, it seems that we cannot even phenomenologically discern *substances* from *events*. Again, I am not denying that substances and events are distinct; rather, I am suggesting that the distinction reflects only different ways of conceptualizing experiences, not differences between the experiences themselves. According to Casati and Varzi (2020), events and substances (they use “objects” to denote substances in their terminology) have different “modes of being”, which are reflected in the predicates we apply to them: we say that a substance *exists*, while an event *occurs*. Furthermore, substances and events bear different relations to space and time. Substances are located at specific spatial points, remaining relatively stable and identifiable over a certain period. By comparison, events are temporal particulars that are more often taken as changes than as persistence. My point is that the features used to distin-

<sup>17</sup> This argument is inspired by Beebe’s unpublished manuscript.

<sup>18</sup> There are some exceptions; e.g., when you see the lightning in the sky, you seem to see an event without any substance participating.

guish between events and substances—either the associated linguistic behaviors or their respective relations to space and time—seem *too abstract* to be reflected in our experience. Some may suggest that the distinction between changing and persisting—properties that can figure in our experience—maps onto the distinction between events and substances and could serve as an experiential marker for the difference between them. But surely we can have experiences of change involving substances, such as a flying frisbee or a chirping bird. We can also experience static events that do not involve obvious changes in properties, such as a boy lying under a tree for a whole afternoon.

### 3.4 Experience of High-level Properties and Phenomenal Contrast

The above argument faces two objections. First, it appears committed to the assumption that phenomenal content cannot represent abstract properties but only properties such as shapes, colors, and movement in visual experiences, or volume, pitch, and timbre in auditory experiences. This so-called *conservative view* of phenomenal content has recently been challenged by philosophers advocating a *liberal view* of phenomenal content. The liberal view posits that phenomenal content can also admit abstract (high-level) properties, such as causation or kinds (Bayne, 2009; Siegel, 2010). Advocates of this view identify experiences of abstract properties by way of the method-of-contrast argument: they describe a pair of contrasting experiences and attribute the relevant difference to phenomenal content representing abstract properties.<sup>19</sup> Along these lines, **Alexa's Transformation** seems to be a kind of contrast argument for the possibility of abstract phenomenal content: the radical change in Alexa's experience can be explained only by the contrasting contents of event-causation and non-event-causation.

A second objection, closely related to the first, argues that while it may be true that we cannot phenomenologically differentiate event-causation from substance-causation in general, we can still phenomenologically distinguish event-causation from agent-causation in particular. Although agent-causation is a type of substance-causation, it possesses distinctive features that are absent in event-causation as well as in other forms of substance-causation; otherwise, we could not explain the radical change of experience involved in **Alexa's Transformation**.

Both objections make use of two premises: first, a contrast argument is a reliable method of identifying high-level or abstract properties in our experience; and second, the scenario of **Alexa's Transformation** serves as a contrast argument identifying the experience of agent-causation or non-event-causation. I dispute the second premise on the grounds that comparing hypothetical experiences that no one has actually had is not a reliable basis for making a contrast argument. More importantly, the question in focus is whether direct experience of event-causation or agent-causation is possible. **Alexa's Transformation** simply stipulates such a possibility without arguing for it.

<sup>19</sup> For instance, Siegel (2009) applies this method to the phenomenology of causation, Bayne (2009) applies it to the phenomenology of kind, and de Vignemont (2020) applies it to the phenomenology of bodily ownership.

A more compelling contrast argument that phenomenal content can include both agent-causation and event-causation would appeal to familiar experiences shared by everyone. Consider cases in which our bodies are caused to move involuntarily by certain mental states, such as when fear triggers a startle response or intense pain causes one to flinch. Call these cases *involuntary mental event-causation*. Two essential features of these cases are noteworthy: first, the bodily movements are commonly taken to be event-caused by the mental states; and second, they exhibit a different phenomenology from the experience of self-as-source. Based on these observations, an argument from phenomenal contrast can be formulated as follows.

(P1) There is a phenomenal difference between experiences of self-as-source and experiences of involuntary mental event-causation.

(P2) The best explanation for P1 is that the two types of experiences involve phenomenal contents representing agent-causation and event-causation respectively.

Therefore,

(C) There are distinct phenomenal contents representing agent-causation and event-causation respectively.

There are indeed differences between the experience of self-as-source and the experience of involuntary mental event-causation. The problem is that these differences can be explained without postulating phenomenal content of agent-causation and event-causation. That is to say, one can (and, I believe, should) accept P1 while denying P2.

In cases of involuntary mental event-causation, we feel as if certain mental states such as fear and pain cause the resultant bodily movements. In contrast, in self-as-source, we do not feel as if our actions are caused by the putative motivational mental events such as beliefs, desires, and intentions. A plausible explanation for this difference is that while we have *salient* feelings of the presence of fear or pain, we normally do not have *salient* feelings of the presence of motivational mental states. There is a grand debate in cognitive phenomenology over whether motivational mental states involve proprietary subjective experience.<sup>20</sup> Fortunately, our purposes do not require us to choose a side. To build up a phenomenal contrast, we need not commit to motivational mental states' having no phenomenology *at all*. We need only commit to their not having *salient* phenomenology as fear and pain do. Do I have any particular feelings about my belief that  $1 + 1 = 2$ ? Does my desire to get published in prestigious journals involve any unique phenomenology? I am not sure. But what I am sure of is that any phenomenal character involved in such mental states would not be as obvious as the phenomenal character of pain or fear.<sup>21</sup> Furthermore, the very

---

<sup>20</sup> For a review of the debates in cognitive phenomenology, see Bayne and Montague (2011).

<sup>21</sup> The case of desires is more complicated. In context, we can distinguish rational desires from appetitive (non-rational) desires. Rational desires are motivated by reasons, whereas appetitive desires are those that befall our bodies, like hunger and thirst. Admittedly, appetitive desires, such as hunger, thirst, or sexual desire, do involve salient bodily feelings (Alvarez, 2009, 73-74). However, when we are acting on appetitive desires, it seems that the stronger the bodily feeling associated with those desires, the more we feel as if our bodies are being pushed, and accordingly, the less prominently the sense of self-as-source is felt.

existence of a debate over whether our motivational mental states have a phenomenology is evidence that whatever phenomenology they do have is not salient.

Another factor contributing to the phenomenal difference between action and involuntary mental event-causation is the sense of control: we feel that we are in charge of our actions, but we do not have the same feeling in involuntary mental event-causation. Again, explaining this difference does not seem to require positing content that represents agent-causation. Studies in cognitive science suggest that this experience can be largely explained by cognitive processes related to action. An influential model to account for this sense of control is the *comparator model*. This model is based on a widely shared theoretical framework for motor control, according to which agents obtain control over their actions through a computational-representational process. Roughly, the control system continually makes predictions about the output of bodily motion while also continually receiving sensory feedback about the output, comparing this feedback with the predictions, and regulating body movements accordingly. The basic hypothesis of this model is that the comparator framework for motor control is also responsible for generating the sense of control—the experience arises when the comparison finds that the actual sensory output matches the predicted output. While the comparator model as an explanation of the sense of control is supported by some empirical evidence, it seems inadequate to accommodate all aspects of the experience.<sup>22</sup>

A more recent and perhaps more promising approach to accounting for the sense of control appeals to metacognition. According to this metacognition approach, the experience arises through second-order cognitive processes and states that monitor first-order cognitive processes underpinning action execution or action selection. For instance, Carruthers (2015) proposes that the sense of control is “the metacognitive monitoring of two cues”: the first cue is the smoothness of action, which refers to the comparison between the predicted output and the actual sensory output, as specified by the comparator framework for motor control. The second cue is the action outcome, which is the comparison between the distal intention and the action effect.<sup>23</sup> Another metacognition account, proposed by Chambon et al. (2014), maintains that a significant component of the sense of control is constituted by a cognitive process that monitors the fluency of action selection. A key study by Wenke et al. (2010) supports this account. In one of their experiments, participants were instructed to press a left or right key depending on whether they saw a left-pointing or right-pointing target arrow. Unbeknownst to them, a subliminal prime arrow was shown before each target arrow, either matching or conflicting with the target. After each keypress, a specific color was shown as an outcome. Importantly, the colors corresponded to the matching or conflicting conditions but did not depend on the primes, the target, or the

<sup>22</sup> For a summary of the empirical evidence for the comparator model, see Bayne and Pacherie (2007). For a recent discussion of the shortcomings of the comparator model, see Haggard (2017) as well as Mylopoulos and Shepherd (2020).

<sup>23</sup> Compare this with Mylopoulos’s account (2017), which also employs higher-order cognitive resources to explain the experience. According to Mylopoulos, the sense of control is constituted by first-personal thoughts based on proximal/executive intentions, and is independent of sensory output. Interestingly, sensory output remains relevant in Mylopoulos’s account because a mismatch between prediction and feedback can disrupt the experience.

actions alone, and thus could not be predicted merely based on the primes, the target, or the actions. This setup helped to factor out the influence of prediction on participants' experience. Interestingly, participants still reported feeling a greater sense of control over the colors corresponding to the matching conditions, suggesting that the ease (fluency) of action selection also influences the sense of control.

While the adequacy of these metacognition accounts remains an open question,<sup>24</sup> my argument only relies on a specific idea they suggest: the sense of control tracks properties involved in the cognitive processes crucial to the planning and execution of action (e.g., the fluency of action selection, the matching of the predicted output and the sensory feedback). This is plausible because the primary function of the sense of control seems to be informing the agent that their action is proceeding well. These cognitive processes, considered as computational-representational, can, in principle, be realized within an event-causal framework. In other words, the properties tracked by the sense of control are compatible with event-causation, which undermines the motivation to postulate additional agent-causation content.<sup>25</sup> Notably, the argument does not assume that the agent is consciously aware of the cognitive processes. If the experience is non-conceptual in the sense that the details of the cognitive processes are not accessible to the agent's awareness, as suggested by Chambon et al. (2014), then it must be neutral on whether action is agent-caused or event-caused.

Let me wrap things up. The experience of self-as-source is *prima facie* incompatible with an event-causal account of action. One response to this challenge, proposed by Horgan and developed by Clarke, is not to construe self-as-source as actually presenting non-event-causation content but only as lacking event-causation content. This move is blocked by PCPN, a compelling principle that links phenomenal difference to phenomenal negation. Nevertheless, to apply PCPN, one must be committed to our having specific phenomenal contents associated respectively with event-causation and non-event-causation. In this section, I have provided reasons to doubt this commitment. I have suggested that people are inclined to mistake the phenomenal difference between involuntary mental event-causation and voluntary action for a phenomenal difference between event-causation and non-event-causation. I have argued that no such difference as the latter is required to explain the former.

## 4 Choice-Making and the Feeling of Detachment

Now let us consider another aspect of the problem of the disappearing agent: the failure of the event-causal framework to account for the agent's role as a choice-maker. This critique has its basis in phenomenology, as many philosophers have argued that the experience of making choices, particularly difficult ones, involves a unique phenomenological feature. During the choice-making process, individuals feel as if they

---

<sup>24</sup> Mylopoulos and Shepherd (2020) offer some reflections on the metacognition approach. For other empirical investigations on the relationship between metacognition and the sense of control, see Metcalfe et al. (2013) and Wen et al. (2023).

<sup>25</sup> A potential concern is that action selection, construed in a specific way, may introduce an independent phenomenological aspect that contradicts event-causation. This concern is addressed in the next section.

stand apart from their mental states in a way that allows them to reflect on, evaluate, or even manipulate these mental states. Thus, individuals experience detachment from their mental states while making a choice. Let us refer to this phenomenological feature as *detachment*. Detachment has been widely acknowledged by both proponents and opponents of CTA.

When an agent reflects on the motives vying to govern his behaviour, he occupies a position of critical detachment from those motives; and when he takes sides with some of those motives, he bolsters them with a force additional to, and hence other than, their own. His role must therefore be played by something other than the motives on which he reflects and with which he takes sides. (Velleman, 1992, 476–77)

The image of the agent directing and governing is, in the first instance, an image of the agent herself standing back from her attitudes and doing the directing and governing. (Bratman, 2005, 33–34)

It is this seeming experience of myself [when making hard choices] as playing a causal role over and above the causal role of my desires and beliefs that suggests I exercised the power of self-determination. (Franklin, 2018, p. 182)

When you deliberate, it is as if there were something over and above all of your desires, something which is *you*, and which *chooses* which desire to act on. (Korsgaard, 1996, p. 100, original emphasis)

Thus, some philosophers are inclined to regard the phenomenology of detachment as incompatible with the event-causal framework. But how? I have argued against the idea that our subjective experience directly exhibits content representing agent- or event-causation. Nevertheless, perhaps rather than directly manifesting agent-causation content, detachment involves a type of content from which we can *infer* the occurrence of agent-causation. Recall the analogy from the perception of age: although a person's age is not directly perceived through visual experience, it can be reasonably inferred from observable facial features. In a similar vein, detachment may provide indirect evidence of agent-causation: if agents can be detached from, or stand apart from, their mental states, manipulating and making decisions about these mental states, then perhaps it can be inferred that they do not consist solely of their mental states, nor is their choosing equivalent to causal interactions among their mental states.

However, it is far from clear that detachment is an accurate description of our experience. Suppose I am making a difficult choice between undertaking a Ph.D. program in philosophy and going to law school for a JD. Would I feel as if I were separate from my attitudes, dispassionately manipulating my beliefs and desires regarding the pros and cons of each option like someone doing mathematical calculations? To me, at least, such an experience seems alien. I find the following a more familiar picture. In making the decision, I am tortured by division within myself and oscillate between options. At times, picturing myself engrossed in interesting philosophical



texts or engaged in philosophical discussions, I am inclined toward pursuing a philosophy Ph.D. Then I remember a blog post by some early-career philosophers complaining about the ultra-competitive job market and their resulting mental suffering. This dampens my enthusiasm for a philosophy career, and I begin to feel more drawn to the law-school option. But then I consider the prospect of poring over monotonous legal cases. After a while, Kant's words about the heavenly stars above and the moral laws within fill my mind and cause goose bumps to appear all over my body...<sup>26</sup> The take-home message is that in making a hard decision, I do not feel detached from my relevant attitudes. Rather, I experience my thinking self as constantly intertwined with them.

Although personal experience alone may not be sufficient to challenge the widely accepted picture of detachment, there are other grounds for questioning its plausibility. Specifically, it leads to conceptual difficulties. Detachment posits that when agents make a choice, they "stand apart" from their attitudes, perhaps comparing or evaluating them. This suggests that, rather than directly choosing one of the actual options at hand, agents are choosing those of their attitudes that favor one option over another. For example, an agent choosing among lunch options is not choosing a burrito over a sandwich but the desire for a burrito over the desire for a sandwich. This is clearly counterintuitive. However, even setting aside the implied overintellectualization, the detachment model raises a more significant issue. All choices, except for random selection, must be based on specific attitudes; if an agent is detached from or stands apart from their attitudes, how can they choose in the first place?

A potential response appeals to higher-level attitudes, such as higher-order desires, to explain the choice between attitudes from which the agent is detached. However, this response leads to a dilemma. At the outset of making a choice, need the agent be detached from their higher-level attitudes? If so, the problem simply recurs at a higher level (how do agents choose between their higher-level attitudes if they are detached from those attitudes?). But if not, that suggests that the agent is intertwined with or represented by their higher-level attitudes, contrary to the detachment picture.

Therefore, the detachment picture leads to conceptual difficulties. Admittedly, this is not a knockdown argument against it as an accurate description of our experience. Something that is problematic on the conceptual level can still be perceivable on the experiential level, like Escher's pictures or Penrose's triangle. But the conceptual difficulty just pointed out at least provides a reason to doubt that the detachment picture accurately represents our experience of making choices.

In denying the detachment picture, I do not mean to deny our ability to reflect on our attitudes or mental states. The detachment picture, I suggest, should be replaced by a more realistic picture that I call *distancing*. We can reflect on certain mental states of ours because we can distance ourselves from them. For example, a person trying to lose weight can reflect on a specific desire to eat a chocolate bar. And a person with preconceptions about certain sexual orientations can reconsider these preconceptions after the experience of interacting with people with different sexual orientations from themselves. However, we should not conflate this ability with

<sup>26</sup> See also Arpaly (2002, Ch. 1) for some vivid examples deployed to argue that "cool hour" deliberation is not the normal case.

detachment. Distancing oneself from a specific mental state need not mean that there is a self detached from one's mental states. Our ability to re-evaluate a specific mental state might simply be due to our already having certain other mental states in operation. For example, we might abandon the desire to eat a chocolate bar because we already hold the belief that it contains more calories than we need. Or we might abandon a bias against members of the LGBTQ+ community because we already appreciate the value of fairness.

## 5 Conclusion

Standard CTA has been criticized for failing to capture the idea that agents actively engage in their actions. It thereby confronts the problem of the disappearing agent. Under the strong reading, this problem is unsurmountable for any version of CTA with an event-causal framework. In this article, I have entertained a way to substantiate the strong reading by reconstructing the problem as a phenomenological challenge. Specifically, certain aspects of the phenomenology of agency seem to involve veridicality conditions that cannot be satisfied by an event-causal account of action. I have reviewed two phenomenological features that are *prima facie* incompatible with the event-causal account, namely self-as-source and detachment. I have then argued that the first can be seen as compatible with an event-causal account of action, while the second is an inaccurate description of our phenomenology.

It is an important question in the philosophy of action whether the common-sense conception that we actively engage in our actions can be captured in an event-causal framework. This article does not provide a positive answer to this question. Nevertheless, it dissolves an important motivation for a negative answer. We cannot justify the claim that an event-causal account fails to accommodate Active Engagement simply by introspecting our phenomenology of agency. Those who want to challenge CTA based on the strong reading of the problem of the disappearing agent need to appeal to resources beyond phenomenology.

**Acknowledgements** I would like to thank Jules Holroyd, Jimmy Lenman, William Hornett, Andrew Christman, Helen Beebee, T. Ryan Byerly, Tony Cheng, Zhiwei Gu, and Shiwei Chen for their insightful comments and suggestions on earlier drafts of this paper. I am also grateful to the anonymous referees for their constructive criticism. My thinking on this project benefited greatly from discussions with participants at the China's Philosophy of Action Summer Workshop in 2021, and the Intelligence, Knowledge and Action Workshop at Tongji University in 2023. This research is supported by "the Fundamental Research Funds for the Central Universities" [2024WKYXQN081].

**Data Availability** Not applicable.

**Code Availability** Not applicable.

**Declarations** The authors have no relevant financial or non-financial interests to disclose.

**Conflicts of Interest/Competing Interests** The authors have no relevant financial or non-financial interests to declare.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adams, F., & Mele, A. (1989). The role of intention in intentional action. *Canadian Journal of Philosophy*, 19(4), 511–531.
- Aguilar, J. (2012). Basic causal deviance, action repertoires, and reliability. *Philosophical Issues*, 22(1), 1–19. <https://doi.org/10.1111/j.1533-6077.2012.00216.x>.
- Arpaly, N. (2002). *Unprincipled virtue: An inquiry into moral agency*. Oxford University Press. <https://doi.org/10.1093/0195152042.001.0001>.
- Audi, R. (1986). Acting for reasons. *The Philosophical Review*, 95(4), 511–546.
- Bayne, T., & Montague, M. (2011). Cognitive phenomenology: An introduction. In T. J. Bayne, & M. Montague (Eds.), *Cognitive phenomenology* (pp. 1–35). Oxford University Press.
- Bayne, T., & Pacherie, E. (2007). Narrators and comparators: The architecture of agentive self-awareness. *Synthese*, 159(3), 475–491. <https://doi.org/10.1007/s11229-007-9239-9>.
- Bishop, J. (1989). *Natural agency: An essay on the causal theory of action*. Cambridge University Press.
- Bishop, J. (2010). Scepticism about natural agency and the causal theory of action. In J. H. Aguilar, & A. A. Buckareff (Eds.), *Causing human actions: New perspectives on the causal theory of action* (pp. 69–83). MIT Press. <https://researchspace.auckland.ac.nz/handle/2292/10115>.
- Bratman, M. (2001). Two problems about human agency. *Proceedings of the Aristotelian Society*, 101(3), 309–326.
- Bratman, M. (2005). Planning agency, autonomous agency. In J. S. Taylor (Ed.), *Personal autonomy* (pp. 35–57). Cambridge University Press.
- Brent, M. (2017). Agent causation as a solution to the problem of action. *Canadian Journal of Philosophy*, 47(5), 656–673.
- Carruthers, G. (2015). A metacognitive model of the feeling of agency over bodily actions. *Psychology of Consciousness: Theory Research and Practice*, 2(3), 210–221.
- Casati, R., & Varzi, A. (2020). Events. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2020). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2020/entries/events/>.
- Chambon, V., Filevich, E., & Haggard, P. (2014). What is the human sense of agency, and is it Metacognitive? In S. M. Fleming, & C. D. Frith (Eds.), *The cognitive neuroscience of metacognition* (pp. 321–342). Springer.
- Clarke, R. (2017). Free will, agent causation, and ‘disappearing agents’. *Noûs*, 53(1), 76–96. <https://doi.org/10.1111/nous.12206>.
- Clarke, R. (2019). Agent causation and the phenomenology of agency. *Pacific Philosophical Quarterly*, 100(3), 747–764. <https://doi.org/10.1111/papq.12275>.
- Davidson, D. (1973). Freedom to act. In T. Honderich (Ed.), *Essays on freedom of action* (1973: 137–56). London: Routledge and Kegan Paul. As reprinted in *Essays on actions and events*, Oxford University Press, 1980, pp. 63–81.
- de Vignemont, F. (2020). What phenomenal contrast for bodily ownership? *Journal of the American Philosophical Association*, 6(1), 117–137. <https://doi.org/10.1017/apa.2019.34>.
- Ekstrom, L. (2000). *Free will: A philosophical study*. Westview.
- Enç, B. (2003). *How we act: Causes, reasons, and intentions*. Oxford University Press.
- Frankfurt, H. (1978). The problem of action. *American Philosophical Quarterly*, 15(2), 157–162.
- Franklin, C. (2016). If anyone should be an agent-causalist, then everyone should be an agent-causalist. *Mind*, 125(500), 1101–1131. <https://doi.org/10.1093/mind/fzv177>.

- Franklin, C. (2018). *A minimal libertarianism: Free will and the promise of reduction*. Oxford University Press.
- Ginet, C. (1990). *On action*. Cambridge University Press.
- Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience*, 18(4), 197–208.
- Himmelreich, J. (forthcoming) (Ed.). The disappearing agent as an exclusion problem. *Inquiry: An Interdisciplinary Journal of Philosophy*.
- Horgan, T. (2007a). Agentive phenomenal intentionality and the limits of introspection. *PSYCHE: An Interdisciplinary Journal of Research on Consciousness*, 13(1), 1–29.
- Horgan, T. (2007b). Mental causation and the agent-exclusion problem. *Erkenntnis*, 67, 183–200.
- Horgan, T. (2011). The phenomenology of agency and freedom: Lessons from introspection and lessons from its limits. *Humana Mente*, 15, 77–97.
- Horgan, T., & Nida-Rümelin, M. (2020). On the satisfaction conditions of agentive phenomenology: A dialogue. In C. Erhard, & T. Keiling (Eds.), *The Routledge handbook of phenomenology of agency* (pp. 264–299). Routledge.
- Horgan, T., Tienson, J. L., & Graham, G. (2003). The phenomenology of first-person agency. In S. Walter, & H. D. Heckmann (Eds.), *Physicalism and mental causation* (pp. 323–340). Imprint Academic.
- Hornsby, J. (2004). Agency and actions. *Royal Institute of Philosophy Supplement*, 55, 1–23.
- Hornsby, J. (2008). Agency and alienation. In de M. Caro, & D. MacArthur (Eds.), *Naturalism in question* (pp. 173–187). Harvard University Press.
- Judisch, N. (2010). Bringing things about. In A. Hazlett (Ed.), *New waves in metaphysics* (pp. 91–110). Palgrave Macmillan UK. [https://doi.org/10.1057/9780230297425\\_6](https://doi.org/10.1057/9780230297425_6).
- Kim, J. (1966). On the psycho-physical identity theory. *American Philosophical Quarterly*, 3(3), 227–235.
- Korsgaard, C. (1996). *The sources of normativity*. Cambridge University Press.
- Lowe, E. J. (2008). *Personal agency: The metaphysics of mind and action*. Oxford University Press.
- Melden, A. I. (1961). *Free action*. Routledge and Kegan Paul.
- Mele, A. R. (2003). *Motivation and agency*. Oxford University Press.
- Metcalfe, J., Eich, T. S., & Miele, D. B. (2013). Metacognition of agency: Proximal action and distal outcome. *Experimental Brain Research*, 229, 485–496.
- Mylopoulos, M. (2017). A cognitive account of agentive awareness. *Mind & Language*, 32(5), 545–563.
- Mylopoulos, M., & Shepherd, J. (2020). Agentive phenomenology. In U. Kriegel (Ed.), *Oxford handbook of the philosophy of consciousness* (pp. 164–187). Oxford University Press.
- Nagel, T. (1986). *The view from nowhere*. Oxford University Press.
- Nida-Rümelin, M. (2007). Doings and subject causation. *Erkenntnis*, 67(2), 255–272. <https://doi.org/10.1007/s10670-007-9064-z>.
- O'Connor, T. (1995). Agent-causation. In T. O'Connor (Ed.), *Agents, causes, and events: Essays on indeterminism and free will* (pp. 173–200). Oxford University Press.
- Pacherie, E. (2012). Action. In K. Frankish, & W. Ramsey (Eds.), *The Cambridge handbook of cognitive science* (pp. 92–111). Cambridge University Press.
- Pereboom, D. (2015). The phenomenology of agency and deterministic agent causation. In H. Pedersen & M. Altman (Eds.), *Horizons of authenticity in phenomenology, existentialism, and moral psychology: Essays in honor of Charles Guignon* (pp. 277–94). *Contributions to phenomenology* (vol. 74). Springer. [https://doi.org/10.1007/978-94-017-9442-8\\_17](https://doi.org/10.1007/978-94-017-9442-8_17).
- Shepherd, J. (2022). Disappearing agents, mental action, rational glue. In M. Brent, & L. M. Titus (Eds.), *Mental action and the conscious mind* (pp. 14–30). Routledge.
- Siegel, S. (2009). The visual experience of causation. *Philosophical Quarterly*, 59(236), 519–540. <https://doi.org/10.1111/j.1467-9213.2008.607.x>.
- Siegel, S. (2010). *The contents of visual experience*. Oxford University Press.
- Steward, H. (1997). *The ontology of mind: Events, processes, and states*. Oxford University Press.
- Steward, H. (2012). *A metaphysics for freedom*. Oxford University Press.
- Taylor, R. (1966). *Action and purpose*. Humanities.
- Thalberg, I. (1984). Do our intentions cause our intentional actions? *American Philosophical Quarterly*, 21(3), 249–260.
- Velleman, D. (1992). What happens when someone acts? *Mind*, 101(403), 461–481.
- Wen, W., Charles, L., & Haggard, P. (2023). Metacognition and sense of agency. *Cognition*, 241, 105622.
- Wenke, D., Fleming, S. M., & Haggard, P. (2010). Subliminal priming of actions influences sense of control over effects of action. *Cognition*, 115(1), 26–38.
- Wilson, J. (2021). *Metaphysical emergence*. Oxford University Press.

- Alvarez, M. (2010). *Kinds of Reasons: An essay in the philosophy of action*. Oxford University Press, USA.
- Bayne, T. (2009). Perception and the reach of phenomenal content. *The Philosophical Quarterly* 59(236), 385-404. <https://doi.org/10.1111/j.1467-9213.2009.631.x>
- Schlosser, M. E. (2010). Agency, ownership, and the standard theory. In A. Buckareff, J. Aguilar, and K. Frankish (Eds.), *New waves in philosophy of action*, pp. 13-31. PAGE NUMBERS IN ROUND BRACKETS Palgrave Macmillan.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.