SOFIA BONICALZI

Università di Pavia
sofia.bonicalzi@gmail.com

DOES REDUCTIVIST EVENT-CAUSAL COMPATIBILISM LEAVE ANYTHING OUT? LYNNE BAKER'S *REFLECTIVE-ENDORSEMENT* AND THE BOUNDS OF THE TRADITIONAL ANALYSES OF MORAL RESPONSIBILITY*

abstract

Promising to be the best companion for scientific naturalism, compatibilism usually espouses a reductivist event-causal background. Lynne Baker challenges this view, arguing that compatibilist moral responsibility also requires an irreducible "first-person perspective". In this paper I will provide some arguments for claiming that (Frankfurt-type) event-causal accounts cannot avoid making reference to some sort of agential properties. In the second part, I will present the proposals formulated by Nelkin and Markosian for defending agent-causation, before returning to the theme with which I began, this time considering Frankfurt's view in the light of Baker's reading.

DOES REDUCTIVIST EVENT-CAUSAL COMPATIBILISM LEAVE ANYTHING OUT?

SOFIA BONICALZI *Università di Pavia*

1. Preliminary Notes

Compatibilism, as I understand it, is a label that characterizes several different views, which share the idea that the truth of determinism is compatible with the existence of free will and/or with the plausibility of moral responsibility attributions. Promising to be the best companion for scientific naturalism, compatibilism – with some notable exceptions (Nelkin 2011; Markosian 1999, 2012; Horgan 2007) – usually espouses an event- (or state-) causal background, in which intentional action is explained in terms of the interaction between different mental states that causally determine one's choices. In her latest book (2013), Lynne Baker claims that moral responsibility (like agency) requires something more than this and, in particular, it requires an irreducible first-person perspective, something that is usually not so welcome in scientific naturalistic views. Since Baker does not want to give up either ("near") naturalism, or some event-causal background, or compatibilism, her proposal sounds especially challenging. The paper is organized as follows. First, in order to carve out a more specific battlefield, I will focus on determinism-friendly accounts originating from Harry Frankfurt's seminal work. My working hypothesis is that Frankfurt-type compatibilism faces some difficulties in explaining how we are in control of our actions and, in particular, what happens when one experiences a clash among different motivational streams. I will try to show that these accounts cannot avoid making reference to some sort of agential properties, and I will mention the proposals formulated by Dana Nelkin and Ned Markosian to defend compatibilist agent-causation. Then – given the dubious implications of these approaches and with some new concepts in place – I will return to the theme with which I began, this time considering Frankfurt's position in the light of Baker's reading.

SOFIA BONICALZI *Università di Pavia*

2. Frankfurt-type Compatibilism and the Disappearing Agent Objection

According to Frankfurt (1988), one acts freely and responsibly, when one acts on a first-order desire that is in accord with a correspondent second-order desire/volition. There is a sense in which, acting as he wants, the *willing addict* (who wants to will to take the drug) is free and responsible for taking the drug, while the *unwilling addict* (who desires to take the drug but does not want to will to take the drug) is not. Moral responsibility is understood in terms of "identification": "Even if the person is not responsible for the fact that the desire occurs, there is an important sense in which he takes responsibility for the fact of having the desire – the fact that the desire is in the fullest sense his, that it constitutes what he really wants – when he identifies himself with it" (Frankfurt 1988, p. 170).

Frankfurt-type compatibilism has been reformulated in several ways, in which different sorts of mental states play the leading role (e.g. Watson 2004, pp. 13-32; Bratman 2001). There are various reasons why these accounts, despite the powerful objections moved by their critics, have a widespread consensus. Much of their appeal resides in the fact that they seem to fit both the *Standard Story* in theory of action and the reductivist view in philosophy of mind, explaining how our actions are up to us (how we can control our conduct) without referring to irreducible agential properties. Mental events play the leading role and, in principle, nothing prevents their reduction to physical states.

However, it is not clear whether these accounts are able to explain control in a proper way. Indeed, it is often held that, both in their libertarian and in their compatibilist interpretations, they are victims of the syndrome of the *disappearing agent*, a version of the more general *luck objection* (Hume 1739; Pereboom 2004, 2012, 2014, 2015; Mele 2006): in the absence of a further explanation of how the choice is up to us, the decision occurs as a result of the causal factors already in place, and there is no way to support ordinary moral responsibility attributions. The fact that the agent might turn out to be a "passive bystander" of a string of mental events represented a serious issue for Frankfurt himself (1988, p. 54). What is doubtful is if identification, or something similar, is sufficient for filling the gap[1]. The problem is that control is hard to reduce to identification with specific mental states: by contrast, it seems that one can control one's choice if one is able to make a decision (at least partially) independently of the motivational force of one's mental states. The situations characterized by the presence of contrasting motives help to stress

---

[1] Bratman expresses a similar concern: "In some cases we suppose, further, that the agent is the source of, determines, directs, governs, the action and is not merely the locus of a series of happenings, of causal pushes and pulls" (2001, p. 311). Velleman explores the idea of the agent as a master of desire in a state or event-causal framework, looking for a kind of mental state that can play a role functionally identical to the role of the agent: "We must therefore look for mental events and states that are functionally identical to the agent, in the sense that they play the causal role that ordinary parlance attributes to him" (Velleman 1992, p. 475).

SOFIA BONICALZI *Università di Pavia*

this point (cfr. Frankfurt 1988, pp. 47-57). The following is a case characterized by opposed first-order desires: Roger wants to climb the *Mount Rushmore National Memorial* but – since there is a fine that discourages people from climbing – Roger opts for avoiding the risk. Now compare Roger with the unwilling addict. The difference rests on the fact that Roger can control his decision, while the unwilling addict does not have this power. The lack of sameness among the desires is not indicative by itself, but only to the extent that it might reveal the practical inability to exercise control over one's desiderative states. Only in such cases the absence of identification undermines moral responsibility. Even in situations of deep *ambivalence* – in which one is divided between opposing second-order desires/ volitions – it is not the lack of identification by itself that does help to illuminate moral responsibility attributions. Consider a less mundane example, a Frankfurt-type version of the story of the *Lady of the Camellias*. Deciding to leave Alfredo under the pressure of his father, Violetta is divided between two opposed second-order desires: she both wants to be moved by the desire to spend her life with Alfredo, and by the desire to help him to have a better life. Being in a condition of ambivalence does not undermine her responsibility. From a phenomenological point of view, Violetta appears to be fully responsible because the decision belongs to her – a reasonable adult woman – independently of her identification with a specific mental state. Moral responsibility attributions depend on the internal structure of her choice in virtue of the fact that she appears to be an agent, who can master different desires, reasons and plans, and whose practical identity goes beyond the sum of her mental states.

However, the idea of an "agent causing an action" does not fit a reductivist event-causal framework, and speaking up for agential properties seems slippery for a variety of reasons. Conceiving the agent as a peculiar substance capable of causally interacting with the physical dimension might not fit the naturalistic vision of the world also in a broadly construed naturalism (De Caro & Voltolini 2010, p. 76).

3. Some Moves towards Compatibilist Agent-causation

One strategy is to replace the event-causal framework with an agent-causal background. The defining claim of agent-causation is that agents are substances capable of causing decisions or intention-formations (Pereboom 2015). The adoption of the agent-causal perspective has traditionally characterized a branch of libertarian views on free will and moral responsibility (O'Connor 2000) but, more recently, it has been also advocated by some compatibilists. For example, Nelkin (2011) and Markosian (1999, 2012) both proposed compatibilist approaches to agent-causation, which deny that it is incompatible with determinism.

Markosian develops a hybrid account, where agent-causation coexists with event-causation inside a materialistic conception. An action is morally free *iff* it is caused by an agent, and the agent is morally responsible *iff* he is the cause of that action.

Admitting *double causation* – according to which the very same event can be produced by two independent factors – an action freely produced by an agent can also be produced in an event-causal way. It might be objected that, if it is only the event-causal stream that is deterministic (while the agent-causal one is indeterministic), the account fails to provide a compatibilist version of agent-causation (Pereboom 2015). Otherwise, if both are deterministic and are causing the very same event, either (a) the physical occurrences characterizing the event-causal stream are not sufficient by themselves and the interaction with the agential causal powers should be explained or (b) the physical occurrences characterizing the event-causal stream are sufficient and the agential causal powers appear to be redundant (and, if the physical realm is complete, there seems to be no reason for admitting extra causal powers [Bennett 2003]). One of the burdens of such a view is that the analysis of the structure of the choice-making process – the core of Frankfurt-type compatibilism – partially loses its centrality. No matter the circumstances of choice, the action is free because an agent produces it (Markosian 2012, p. 384). Then Markosian – with the questionable assumption that, if the action is morally wrong, then it has to be morally free – has to admit that also a brainwashed individual (like Patriot Kid, who shoots the president after being kidnapped and manipulated by Martians [Markosian 1999, p. 272]) is morally responsible.

Nelkin instead provides a unified account, in which the only form of causation that exists is substance causation (Lowe 2008), whose effects are determined: given the kind of substance the agents are, and the circumstances in which the action occurs, the choice is deterministically produced. Nelkin adopts a distinction made by O'Connor (2000; cfr. Dretske 1993) between *structuring* and *triggering* causes. While reasons are the structuring causes of a choice (structuring one's propensities), the agent, with his specific causal power, is the triggering cause that settles the final decision. In our story, Violetta, with her specific causal power, makes a choice, which turns out to be determined. A source of doubt is that – once one's propensities have been already structured – it is not clear how the power to settle the (determined) final decision is to be understood in a way that might preserve one's ability to control one's own conduct. Despite some obscurities, deterministic agent-causation might represent a promising path. However, for many – at least without further clarifications – the assumption that the agents cause events remains controversial or even unintelligible, and turns out to be a sort of *ignotum per ignotius* explanation.

4.
Lynne Baker's
Reflective-
Endorsement

A different perspective is sketched by Lynne Baker. In her work, Baker addresses directly the problem of moral responsibility recruiting Frankfurt's event-causal framework, which might be revised in the light of an explicit reference to the ro-bust form of the *first-person perspective on agency*, defined as "the capacity to think of oneself, conceived in the first person, as the object of one's thought" (2013, p. XIX),

and intended as incompatible with a third-person ontology[2]. Why is this "defining characteristic of persons" (2013, p. 201) – who "can consider the reasons" they "have and choose to act on them" (2013, p. 202) – supposed to improve Frankfurt-type approaches?

The most peculiar aspect of Frankfurt's hierarchical view consisted in the idea that, in order to be morally responsible, one should be able to conceive the mental states leading to action as one's own. In Baker's *Reflective-Endorsement*, it is this essential capacity that gives people the limited amount of control – the ability to consider the reasons we have and to act on them – that might save the day for compatibilism[3]. More precisely, an "agent is morally responsible for an action if he endorses the beliefs and the desires on which he acts: When he affirms them as his own [...], he is morally responsible for acting on them" (2013, p. 205). According to Baker, the appeal to the first-person perspective is not to be intended merely as a reference to the practical unity of the subject, but implies an ontological commitment. As mentioned earlier, one intriguing aspect of Baker's proposal is that it is committed to preserve an event-causal framework. But how is the concept of an "event" to be understood? Following Kim, Baker interprets an event as an object's having a property at a time (2007, p. 97). What should be abandoned is rather the reductivist spirit according to which conscious mental events are reducible to physical states. In virtue of being (emergent) upper level properties-instances, mental states are irreducible to lower-level physical properties-instances, turning out to be independently causally efficacious[4]. The relation between the two orders is conceived in terms of *constitution*: given certain *favorable circumstances*, the higher-level properties-instances are constituted by, but not reducible to, the lower level ones, as a cat

---

2   "Frankfurt, Velleman, and Bratman [...] all speak of an agent's reflective participation in her action as if reflective participation [...] is compatible with a third-person ontology. Many philosophers do not acknowledge that the first-person perspective presents a problem for scientific naturalism" (Baker 2013, p. XVII).

3   The *Reflective Endorsement* view is articulated as follows.

(RE) A person S is morally responsible for a choice or action X if X occurs and:

(1)   S wills X,

(2)   S wants that she* will X [i.e., S wants to will X],

(3)   S wills X because she* wants to will X, and

(4)   S would still have wanted to will X even if she had known the provenance of her* wanting to will X.

Where the fourth condition specifies that the agent would not repudiate her desires given that she is aware of their provenance, and the "*" identifies the first-person perspective (Baker 2013, p. 204).

4   Baker vindicates *commonsense causation* (as making something happen, giving rise to something): "An object x (or a property instance) has causal powers if and only if x has a property F in virtue of which x has effects" (2007, p. 98. See also Baker 2011, pp. 12-13). About Baker's emergentism, see instead Baker 2013, p. 220; 2007, p. 237.

SOFIA BONICALZI *Università di Pavia*

is constituted by, but not reducible to, the sum of its particles[5]: "The Constitution View, applied to property-instances, allows intentional phenomena to have causal efficacy" (Baker 2011, p. 13). For making sense of the moral realm, some first-person properties should be admitted in our ontology: "Property P is a first-person property if either (1) P entails that whatever exemplifies it has the capacity to interact consciously and intentionally with the environment and/or (2) P entails that whatever exemplifies it can conceive of herself as herself* in the first-person" (Baker 2013, p. 172).

Without considering the traditional objections towards the anti-reductivist program in general, a doubt I wish to explore in the remaining part of the paper regards constitution and its dependence on some favourable circumstances. Irreducible (and causally efficacious) emergent properties are produced by their subatomic constituents, given the presence of the relevant circumstances[6]. Differently from *supervenience* (which is necessary and independent of contextual factors), constitution occurs only if the microphysical constituents are accompanied by the relevant circumstances, so that "although a constituting property-instance does not supervene on its constituting property-instances, it may supervene ultimately on its subatomic constituters *together with* the microphysical supervenience base of all the circumstances in which the instance of the constitution relation obtains" (Baker 2013, pp. 219-220). Since Baker's near naturalism leaves the door open for the truth of the *causal-closure thesis* (*ibidem*), the emergent properties are constituted by their microphysical particles plus the relevant circumstances that, in turn, are also constituted by their microphysical particles. Then (even though that particular lower-level event does not necessitate that particular higher-level event), one might object that the upper-level properties and, in particular, the first-person perspective – the *locus* where the non-biological (Baker 2000, p. 17) discontinuity between human and non-human animals takes place – turn out to be a practical (epiphenomenal?) stance with no really independent causal powers, a lens through which one regards oneself as a unity, but without having a grasp of the ontological reality.

My last concern regards the kind of moral responsibility that is in question. The direction taken by Baker to escape the *disappearing agent objection* is quite promising: the implicit reference to the first-person point of view –

5 About mental causation and constitution, see Baker 1993, 2007, 2013. For a different account of constitution ("*the made up of* relation"), inside a non-reductivist framework, see Pereboom 2011, pp. 135-141.
6 "Whether or not constitution obtains depends in large measure on the circumstances" (Baker 2013, p. 209).

which is hidden in Frankfurt's approach – is thus made explicit and it is now possible to account for cases of opposing desires and ambivalence, in which one is in control of one's choice because one refers the opposing mental states to oneself. Nevertheless, even incompatibilists usually do not deny that, also given the truth of determinism, one can endorse one's beliefs and desires, or think about the origins of one's mental states, thus forming *a sort of* first-person perspective, and having the impression of being able to shape the causes of one's choices. However, nothing proves that this picture, which fits a certain phenomenology of agency, is not a *post-factum* illusory reconstruction (despite Baker's denial: "The first-person perspective cannot be acquired by neural manipulation" [2013, p. 202; see also Baker 2006]), or a "center of narrative gravity", to use Daniel Dennett's words (1992). To dismiss these worries, Baker claims that her core concerns diverge, for example, from those that inspire the *theories of confabulation* in cognitive sciences (Carruthers 2013. See also e.g. Wegner 2002; Preston & Wegner 2005; Marraffa & Paternoster 2013), which are mainly focused on the idea that, given for example the *opacity* of introspection (Carruthers 2011), we might be mistaken "about the sources of our first-order beliefs" (Baker 2013, p. 64) or about the content of our inner life. Baker's theory rather concerns the question: "under what conditions can we have beliefs about our beliefs at all?" (2013, p. 64): having a robust first-person perspective, or "conceiving of oneself as having a perspective" (2013, p. 82) is meant to be the basic requisite for having a inner life, something that cannot be understood in terms of a misleading rationalization and self-ascription of mental states. Yet, does my awareness of my inner life – no matter the possible lack of insight into my first-order mental states – represent a strong basis for moral responsibility attributions? The problem with compatibilism does not seem to be that one might be unable to conceive of oneself as oneself. The limit is rather that – once accepted that our choices are determined by factors that are beyond us – we could hardly make sense of the concept of *accountability*[7] that, for many, is what moral responsibility is supposed to be. Nevertheless, even though it is unlikely that Baker's account proves successful against traditional incompatibilist worries, it opens a thought-provoking line for those who share compatibilist intuitions and, more generally, for those – including myself – who are inclined to think that moral responsibility has (much) to do with identification with motives mediated through practical reasoning.

---

7   At least for certain interpretations, the idea that one deserves to be blamed and praised in virtue of the choice one made (cfr. Watson 2004, pp. 260-288; Pereboom 2001, p. XX).

REFERENCES

Baker, L.R. (1993), "Metaphysics and Mental Causation", in J. Heil & A. Mele (eds.), *Mental Causation*, Clarendon Press, Oxford, pp. 75-96;

Baker, L.R. (2000), *Persons and Bodies*, Cambridge University Press, Cambridge;

Baker, L.R. (2006), "Moral Responsibility Without Libertarianism", *Noûs*, 42, pp. 307-330;

Baker, L.R. (2007), *The Metaphysics of Everyday Life*, Cambridge University Press, New York;

Baker, L.R. (2011), "First-Personal Aspects of Agency", *Metaphilosophy*, 42(1-2), pp. 1-16;

Baker, L.R. (2013), *Naturalism and the First-Person Perspective*, Oxford University Press, New York;

Bennett, K. (2003), "Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract It", *Noûs*, 37(3), pp. 471-497;

Bratman, M. (2001), "Two Problems About Human Agency", *Proceedings of the Aristotelian Society*, New Series, 101, pp. 309-326;

Carruthers, P. (2011), *The Opacity of Mind: An Integrative Theory of Self-Knowledge*, Oxford University Press, Oxford;

Carruthers, P. (2013), "Mindreading the Self", in S. Baron-Cohen, H. Tager-Flusberg & M. Lombardo (eds.), *Understanding Other Mind: Perspectives From Developmental Social Neuroscience*, Third Edition, Oxford University Press, New York, pp. 467-486;

De Caro, M. & Voltolini, A. (2010), "Is Liberal Naturalism Possible?", in M. De Caro & D. Macarthur (eds.), *Naturalism and Normativity*, Columbia University Press, New York, pp. 69-86;

Dennett, D.C. (1992) "The Self as a Center of Narrative Gravity", in F.S. Kessel, P.M. Kohl & D.L. Johnson (eds.), *Self and Consciousness: Multiple Perspectives*, Lawrence Erlbaum Associates, Hillsdale, NY, pp. 103-115;

Dretske, F. (1993), "Mental Events as Structuring Causes of Behavior", in J. Heil & A. Mele (eds.), *Mental Causation*, Clarendon Press, Oxford, pp. 121-136;

Frankfurt, H. (1988), *The Importance of What We Care About: Philosophical Essays*, Cambridge University Press, New York;

Frankfurt, H. (1999), *Necessity, Volition, and Love*, Cambridge University Press, New York;

Horgan, T. (2007), "Mental Causation and the Agent-Exclusion Problem", *Erkenntnis*, 67, pp. 183-200;

Hume, D. (1739/1888), *A Treatise of Human Nature*, L.A. Selby-Bigge (ed.), Oxford University Press, Oxford;

Lowe, E.J. (2008), *Personal Agency: The Metaphysics of Mind and Action*, Oxford University Press, Oxford;

Markosian, N. (1999), "A Compatibilist View of the Theory of Agent Causation", *Pacific Philosophical Quarterly*, 80, pp. 257-277;

Markosian, N. (2012), "Agent Causation as the Solution to All the Compatibilist's Problems", *Philosophical Studies*, 157, pp. 383-398;

Marraffa, M. & Paternoster, A. (2013), *Sentirsi esistere. Inconscio, coscienza, autocoscienza*, Laterza, Roma-Bari;

Mele, A. (2006), *Free Will and Luck*, Oxford University Press, New York;

Nelkin, D. (2011), *Making Sense of Freedom and Responsibility*, Oxford University Press, New York;

O'Connor, T. (2000), *Persons and Causes: The Metaphysics of Free Will*, Oxford University Press, New York;

Pereboom, D. (2004), "Is Our Conception of Agent-Causation Coherent?", *Philosophical Topics*, 32, pp. 275-286;

Pereboom, D. (2011), *Consciousness and the Prospects of Physicalism*, Oxford University Press, New York;

Pereboom, D. (2012), "The Disappearing Agent Objection to Event-Causal Libertarianism", *Philosophical Studies*, 169, pp. 59-69;

Pereboom, D. (2014), *Free Will, Agency and Meaning in Life*, Oxford University Press, Oxford;

Pereboom, D. (forthcoming 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*, Springer, New York;

Preston, J. & Wegner, D.M. (2005), "Ideal Agency: On Perceiving the Self as an Origin of Action", in A. Tesser, J. Wood & D. Stapel (eds.), *On Building, Defending, and Regulating the Self*, Psychology Press, Philadelphia, pp. 103-125;

Velleman, J.D. (1992), "What Happens When Someone Acts", *Mind*, New Series, 101(403), pp. 461-481;

Watson, G. (2004), *Agency and Answerability: Selected Essays*, Oxford University Press, Oxford;

Wegner, D.M. (2002), *The Illusion of Conscious Will*, MIT Press, Cambridge (MA).