Erkenntnis (2022) 87:1889–1906
https://doi.org/10.1007/s10670-020-00281-w

ORIGINAL RESEARCH

Free Will, Control, and the Possibility to do Otherwise from a Causal Modeler's Perspective

Alexander Gebharter[1] · Maria Sekatskaya[2] · Gerhard Schurz[2]

Received: 26 April 2019 / Accepted: 27 May 2020 / Published online: 20 June 2020

Abstract
Strong notions of free will are closely connected to the possibility to do otherwise as well as to an agent's ability to causally influence her environment via her decisions controlling her actions. In this paper we employ techniques from the causal modeling literature to investigate whether a notion of free will subscribing to one or both of these requirements is compatible with naturalistic views of the world such as non-reductive physicalism to the background of determinism and indeterminism. We argue that from a causal modeler's perspective the only possibility to get both requirements consists in subscribing to reductive physicalism and indeterminism.

1 Introduction

Free will is still one of the most heavily discussed issues in philosophy. Most philosophers (except free will sceptics, hard determinists, and hard incompatibilists) as well as most non-philosophers (cf. Nahmias et al. 2005; Nahmias 2011; Sarkissian et al. 2010) believe that we have free will. The disagreement begins with the discussion about which analysis of free will gets things right. Many such analyses have been proposed in the philosophical literature and each of them comes with its own merits and problems. Philosophical theories cover a multitude of notions of free will ranging from libertarian ones, requiring an agent to be able to act one way or another while holding the laws of nature and the past fixed while at the same time providing the agent with ultimate control over her actions (cf. Kane 1994, 1996; van Inwagen 1975, 1983), to compatibilist and semi-compatibilist notions, arguing that free will is compatible with there being only one course of action open to the agent (Dennett 1984; Frankfurt 1969).

✉ Alexander Gebharter
alexander.gebharter@gmail.com

[1] Department of Theoretical Philosophy, University of Groningen, Oude Boteringestraat 52, 9712 GL Groningen, The Netherlands

[2] Düsseldorf Center for Logic and Philosophy of Science (DCLPS), University of Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany

There are different taxonomies of positions in the free will debate, usually based on possible combinations of different answers to the following questions (see, e.g., Elzein and Pernu 2017, p. 227; Pereboom 2001, p. xix; Strawson 1986, p. 6): Does free will require alternative possibilities? Are alternative possibilities compatible with physical determinism? Does physical determinism obtain? Does free will require ultimate control? Is this kind of control compatible with physical determinism and/or indeterminism? In this paper we distinguish between strong notions of free will which assert that certain metaphysical truths about causality, agency, and control must obtain in the actual world if we are to have free will, and weak notions which claim that free will is compatible with whatever turns out to be the case according to our best physical theories. This division cuts across the traditional compatibilist versus incompatibilist (libertarian and hard incompatibilist) spectrum, and reflects the fact that most contemporary compatibilists do not argue that physical determinism is either true or necessary for free will. Examples of strong notions can be found in libertarian (e.g., Kane 1994; O'Connor 2000), hard incompatibist (e.g., Pereboom 2001) and strong compatibilist (soft-determinist) theories (e.g., Hobart 1934). Weak notions of free will include, among others, conditional analysis, mesh theories (e.g., Frankfurt 1971), and the Strawsonian approach.

Whatever one's opinions concerning determinism and indeterminism, if one is to have a complete theory of free will, one must explain the causal mechanisms underlying it. Having control over one's actions and thoughts, for example, is clearly a causal notion (cf. Fischer and Ravizza 1998). If freedom requires the ability to choose between different future paths of action, then an agent's decisions seem to be required to make a (causal) difference for which future events come about. Compatibilist dispositional analysis of abilities must provide a description of conditions under which an agent is disposed to perform an act, i.e., to cause an event to happen (Lewis 1997; Vihvelin 2004). Similar considerations apply to libertarian accounts (see, e.g., Clarke 2003) which have to explain how exactly an agent's actions and their sources fit into the causal structure of the world, which role they play therein, and so on and so forth. Though this close connection to causation exists, free will is typically not discussed to the background of a specific approach to causation and most of the debate relies on a rather informal understanding of that concept. But as Hitchcock (2012) demonstrated, the outcomes of philosophical debates involving causal concepts might be heavily influenced by the specific understanding of causation endorsed.

In this paper we investigate strong notions of free will as characterized above to the background of the causal interpretation of the Bayesian network machinery (Pearl 1988) as presented in (Spirtes et al. 1993). Causal Bayes nets are widely appreciated (in philosophy, computer science, and psychology), since they seem to give us the best grasp on causation we have so far. Given that the causal Markov condition is satisfied (for details, see Sect. 4), the framework allows for testing causal hypotheses on the basis of empirical data and for developing powerful algorithmic procedures for uncovering causal structure (ibid). Moreover, as elaborated in (Gebharter 2017b; Schurz and Gebharter 2016), the causal Markov condition provides the best explanation for certain empirical phenomena and if enriched by weak additional assumptions (such as a weak version of faithfulness), it generates novel

predictions by which it can be tested. Thus, the approach seems to satisfy standards of successful theories of the sciences. Our hope is that the framework can also be used as a tool for shedding new light on free will. Since there is a whole plethora of different notions of free will, we will not be able to cover all of them in this paper. To start this project, we focus on strong incompatibilist notions of freedom involving a certain kind of control and the possibility to do otherwise and combine these notions with certain metaphysical assumptions. Note that reconstructing these notions within a causal Bayes net setting will involve some idealization and abstraction. In the end, we will get precision at the cost of losing some details and nuances.

We aim at investigating free will from a naturalistic angle according to which everything going on is based on facts of physics and obeys the laws of nature. In particular, we will focus on positions according to which an agent's free will, decisions, reasons, etc. at least supervene on the physical. It is worth noting that most participants in the contemporary free will debate strive to formulate and defend theories compatible with such a naturalistic worldview.[1] This sets the stage for two ontological positions to whose background we will assess strong incompatibilist notions of free will: non-reductive and reductive physicalism. But note that the project of this paper is not to argue for non-reductive or reductive physicalism or for any particular theory of free will. Instead, we are interested in the consequences of analyzing strong notions of free will to the background of a causal modeler's understanding of causation. We show that this analysis produces unexpected results concerning the possibility of reconciling different demands that certain theories of free will make with different combinations of four broad ontological positions: determinism, indeterminism, and non-reductive and reductive physicalism.

The paper is structured as follows: In Sect. 2 we say more about the strong notions of free will to be investigated in subsequent sections. In Sect. 3 we briefly discuss non-reductive physicalism, reductive physicalism (reductionism for short), physical determinism, and physical indeterminism. In Sect. 4 we present the basics of the Bayes net formalism and the causal interpretation of Bayes nets used throughout the rest of the paper. In Sect. 5 we translate the commitments of strong incompatibilist notions of free will introduced in Sect. 2 and the four positions introduced in Sect. 3 into causal Bayes net terminology. We develop two simple models on whose basis we investigate the compatibility of these commitments with different combinations of these four positions. We conclude in Sect. 6.

[1] Compatibilists, unsurprisingly, are usually in favor of naturalist accounts of free will. Hard incompatibilists often claim that the reason why they assert that people do not have free will is precisely the impossibility to reconcile the conditions that agents must meet in order to have free will with what science tells us about the world (cf. Pereboom 2001). Event-causal libertarians (cf. Ekstrom 2003; Kane 1996, 2007) and non-causal libertarians (cf. Ginet 1990; Pink 2016) claim that their accounts are naturalistic. Departing from naturalism has become a rather unusual move, made by some agent-causal libertarians (Chisholm 1964/2015, 1976) and by the so-called mysterians (van Inwagen 1983, 2000; Shabo 2011).

2 Free Will, Control, and the Possibility to do Otherwise

To start the project outlined in Sect. 1, we focus on strong incompatibilist notions of free will. Such notions presuppose an agent's ability to do otherwise. According to this requirement, an agent is free only if she could have done (or decided to do) otherwise, given that the past and the laws are fixed. Proponents of strong incompatibilist notions argue that having this ability is incompatible with determinism (cf. van Inwagen 1983, 2008). To avoid this kind of worry, some compatibilists have proposed definitions of free will that do not rely on the possibility to do otherwise (see, e.g., Frankfurt 1969; Watson 1975; Dennett 1984). Classical compatibilists, on the other hand, redefine the ability to do otherwise in conditional or dispositional terms (Hume 1748/1999; Saunders 1968; Lewis 1997; Vihvelin 2004; Fara 2008). We do not consider the latter approaches in this paper, because they rely on weak notions of free will which deny that an agent's abilities have causal status different from that of other physically realized dispositions. According to a strong compatibilist understanding of free will, which is also known as "soft determinism", free will demands physical determinism and is inconceivable without it, as Hobart claims in his famous paper (Hobart 1934). This position was quite popular in the first decades of the twentieth century, when physics taught that the world is governed by deterministic laws (Moore 1912; Ayer 1954). We do not discuss soft determinism in detail in this paper; however, in Sect. 5 we show that at least one of the models which we propose can capture some strong compatibilist intuitions.

Incompatibilists have traditionally claimed that a conditional analysis of free will, even if it can be spelled out without contradictions, is nothing over and above a trick. According to the incompatibilists, the analysis of "could have done otherwise" along the lines of "would have done otherwise if" misrepresents our modal intuitions. Incompatibilists claim that their understanding of free will captures the intuitive meaning that non-philosophers have in mind when asserting that someone could have done otherwise.[2] There seems to be nothing "iffy" about abilities, they are categorical (Austin 1979): If an agent is free, then, given the same circumstances (i.e., the past together with the laws of nature), she could have decided to do otherwise (see, among many others, van Inwagen 1983; Ginet 1990; Kane 1996; Ekstrom 1998). Let us label this strong incompatibilist requirement for an agent's free will as follows:

(PDO) The agent could have decided to do otherwise (given the actual past together with the laws of nature).

Another concept strong notions of free will require is the concept of control. According to the leeway libertarians, the relevant control is understood as

[2] Whether this is indeed what non-philosophers have in mind has become a matter of controversy in recent years (Nahmias et al. 2005; Nichols 2011; Sarkissian et al. 2010). However, since we are exploring conceptual possibilities of combining different philosophical theses in this paper, we do not take a stance on discussions in experimental philosophy about how non-philosophers understand free will.

including the ability to do otherwise, while the sourcehood libertarians lay emphasis on whether an agent is the relevant source of her action, even if she could not have done otherwise in some particular situation.[3] In this paper, we have a certain kind of control in mind: For an agent to be called free in this sense, she must be able to control at least some of her actions on the basis of her decisions, meaning that at least some of the agent's decisions must have a probabilistic causal impact on her actions. In other words: The agent's decisions must be able to make a probabilistic causal difference. Note that this is a very weak sense of control that covers strict as well as probabilistic versions of control. Without this assumption, **(PDO)** would be more or less worthless. Even if an agent would be able to decide to do otherwise, different decisions would never lead to different consequences and she would have no possibility to willingly influence the future whatsoever. To capture the idea that the possibility to do otherwise can be exercised by the agent as a continuation of the actual world, some authors refer to this possibility as "actualist" (cf. Elzein and Pernu 2017). Let us also label this second requirement for a strong notion of an agent's free will:

(CTRL) The agent's decisions have a causal impact on her environment (including her actions and attempts to act).

Finally, note that **(PDO)** and **(CTRL)** are logically independent, meaning that someone can consistently subscribe to one but not to the other. An agent could have the possibility to decide otherwise even if it would make no difference for the outcome if she had done so. And, vice versa, an agent's decisions could be (causal) difference makers for parts of her environment even if she could not have decided to do otherwise. For someone subscribing to a strong incompatibilist notion of free will, both alternatives seem unsatisfying. If one wants to have a notion of freedom in the sense that her decisions could have been different in an incompatibilist sense of "could", while at the same time actually making a difference to the world, one needs to subscribe to both, **(PDO)** and **(CTRL)**.

3 Non-reductive Physicalism, Reductionism, Determinism, and Indeterminism

In this paper we combine strong notions of free will with reductive and non-reductive variants of physicalism. We start with non-reductive physicalism. Non-reductive physicalism is understood as a position that grants a certain amount of autonomy to higher-level phenomena. Therefore, it might look especially advantageous for libertarianism.

[3] For a difference between leeway and sourcehood libertarians see (Pereboom 2001, pp. 1–6). However, most libertarians agree that both conditions should be met at least sometimes: If the agent is the ultimate source of her actions, then she can, at least sometimes, behave one way or another. And if it is up to the agent to choose her course of actions, then she is, in some sense, the source of her actions (cf. Kane 1996).

Non-reductive physicalism subscribes to the following three theses (cf. Kim 2005): Higher-level properties are non-identical and not reducible to fundamental physical properties.[4] Higher-level states (i.e., instantiated higher-level properties) supervene on physical states (i.e., instantiated physical properties). And, finally, higher-level properties sometimes play a role in bringing about lower-level states. The first of these three theses drives an ontological wedge between higher levels and the fundamental physical level. It makes room for the view that properties relevant for free will such as making a decision or having a reason populate a level that is ontologically different from the fundamental physical level. Let us call this level the agential level. The second of the three theses establishes a minimal connection between higher levels and the fundamental physical level: Every change at a higher level is necessarily associated with changes at the fundamental physical level. The third thesis finally reflects the commitment that higher-level properties are at least sometimes causally autonomous. Specified to the agential level (populated by an agent's decisions, reasons, and so on) these three theses can be formulated as follows:

(ONI) Agential properties (e.g., making a decision) are on a higher ontological level than (and, thus, non-identical with) physical properties.

(SUP) Agential states supervene on physical states.

(CE) Agential properties are at least sometimes causally relevant for bringing about states of the agent at the physical level.

Reductionism, on the other hand, basically denies the non-identity of higher-level properties with fundamental physical properties. According to this view, all higher-level properties are reducible to fundamental physics. Specified to the agential level this amounts to:[5]

(RED) Agential properties (e.g., making a decision) are identical with (and, thus, reducible to) physical properties.

Non-reductive physicalism as well as reductionism can be combined with determinism or indeterminism:

(DET) Given a specified way things are at a time t, the way things go thereafter is fixed by the laws of nature (cf. Hoefer 2016).

(IND) It is not the case that given a specified way things are at a time t, the way things go thereafter is fixed by the laws of nature.

[4] Throughout the paper we assume that it is intuitively clear enough what counts as a higher-level property, what counts as a fundamental physical property, and what it roughly means that they might populate different ontological levels. Since there is no consensus in the philosophical literature on levels (see, e.g., Eronen and Brooks 2018) we stay neutral on the specific theory of levels in this paper.

[5] Note that since agential properties are identified with properties at the physical level in reductionism, this position is committed to **(SUP)** as well.

In the following, we understand the two latter positions as views about the fundamental physical level. At first glance, it seems that non-reductive physicalism as well as reductionism allow for a strong incompatibilist notion of free will subscribing to both **(PDO)** and **(CTRL)** if the world is indeterministic. This combination of claims constitutes the core of all libertarian accounts. And if determinism is true, then it seems plausible that free will is at least compatible with non-reductive physicalism: Even if the fundamental physical level is fully determined, higher-level phenomena might be multiply realizable by lower-level phenomena and, thus, there might be room for the possibility to do otherwise at the agential level. List (2014) argues that the possibility to do otherwise at the agential level gives us **(PDO)** without further libertarian commitments.[6] Therefore, both traditional libertarian as well as agential level libertarian accounts, seem to demand non-reductive physicalism in order to satisfy both **(PDO)** and **(CTRL)**. Incompatibilists argue that being able to do otherwise is incompatible with physical determinism (cf. van Inwagen 1983, 2008; see Hausmann 2019 for a recent evaluation of van Inwagen's argument). Taking all this into account, we arrive at a first wild guess about the possibility of **(PDO)** and **(CTRL)** summarized in the following table:

	Determinism	Indeterminism
Non-reductive physicalism	**(PDO) + (CTRL)**	**(PDO) + (CTRL)**
Reductive physicalism	×	**(PDO) + (CTRL)**

Note that we do not claim that the table's cells should actually be filled in as above. The point is rather to illustrate that how exactly they should be filled in is not entirely clear only on the basis of **(PDO)**, **(CTRL)**, **(ONI)**, **(SUP)**, **(CE)**, **(RED)**, **(DET)**, and **(IND)**. This is why we speak of a first wild guess above. How to fill in these cells crucially depends on the specific understanding of causation endorsed. In Sect. 5 we will build two causal Bayes net models and translate **(PDO)**, **(CTRL)**, **(ONI)**, **(SUP)**, **(CE)**, **(RED)**, **(DET)**, and **(IND)** into the language of these models. We will then use the causal Bayes net machinery as a tool for determining how to fill in the cells of the table. If one is sympathetic to the first wild guess provided above, one might find these results quite surprising. It will turn out that non-reductive physicalism cannot account for **(CTRL)**, regardless of whether determinism or indeterminism is true, and that reductionism actually fares better than non-reductive physicalism in both cases. Even more surprisingly, it will turn out that libertarians who accept **(PDO)** and **(IND)** by definition, and are often sympathetic to both

[6] For objections, see (Elzein and Pernu 2017; Gebharter ms).

(CTRL) and non-reductive accounts of consciousness, which would imply **(ONI)**, cannot have all four of them.[7] But first, we will briefly introduce the basics of the causal Bayes net formalism required for this analysis.

4 Causal Bayes Nets

Bayes nets can be used to represent all kinds of relations and dependencies that have just the right formal properties. In particular, they must conform to the Markov condition (Spirtes et al. 1993, p. 33). There are several versions of the Markov condition and how to best present it depends to some extent on the purpose one has in mind. For this paper, we choose the following version, where V is a set of variables $X_1, \ldots, X_n$ representing states or properties, E is a set of arrows ($\longrightarrow$) connecting pairs of variables in V, P is a probability distribution over V, and $Par(X_i)$, the set of X_i's parents, is defined as the set of all $X_j \in V$ with $X_j \longrightarrow X_i$:

(Markov condition) $\langle V, E, P\rangle$ satisfies the Markov condition if and only if $G = \langle V, E\rangle$ and P conform to

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | Par(X_i)). \tag{1}$$

If satisfied, the Markov condition guarantees that every variable $X_i \in V$ becomes independent of its non-descendants (i.e., the variables X_j not connected via a path $X_i \longrightarrow \cdots \longrightarrow X_j$ to X_i) conditional on X_i's direct ancestors (i.e., the variables in $Par(X_i)$). Now one kind of relation that conforms to Eq. (1) and, thus, could be represented by a Bayes net's arrows, is direct causal dependence (Spirtes et al. 1993). A Bayes net in which some (or all) arrows are causally interpreted is called a causal Bayes net (CBN). Other relations that seem to conform to Eq. (1) are, for example, supervenience and constitution (Gebharter 2017a, c). Recent work by Schaffer (2016) suggests that also the grounding relation might be representable by the arrows of a Bayes net. In its causal interpretation, the Markov condition establishes a connection between causal structures and probability distributions. In particular, it

[7] Robert Kane, to give just one example, subscribes to this position. Having argued at length that for an agent to have free will she must (i) have alternative possibilities open to her, and (ii) exercise control (in his theory, it is "plural (or dual) rational control"), he writes that the goal of his account "is not to eliminate all mystery from free will. Rather, it is to eliminate mysteries that are created by taking a distinctively libertarian view of free will—as opposed to mysteries that confront everyone, no matter what their position on free will [...]. Indeterministic efforts are mysterious because they partake of several deep cosmological problems that are problems for everyone, not just libertarians. One of these problems is 'the mind/body problem,' including at its core the 'problem of consciousness': How can thoughts, perceptions, and other conscious experiences—including efforts of will—be brain processes? But this is a problem whether you are a compatibilist or incompatibilist, libertarian or nonlibertarian. It is no less mysterious how neural firings in the brain could be conscious mental events if they are determined than if they are undetermined, or if they involve undetermined chaotic processes than if they do not" (Kane 1996, p. 159).

guarantees that every dependence between variables in V can be explained by some causal connection between these variables. Among many other consequences, the Markov condition implies Reichenbach's (1956) insights that common causes screen off their effects and that direct causes screen off effects from their indirect causes.

Before we go on, let us briefly explain the status of the Markov condition for our understanding of causation. Following (Spirtes et al. 1993), we understand causation as one of those relations that conform to the Markov factorization. Note that, according to this understanding, causation is only implicitly characterized and not explicitly defined. Causation understood in this manner has, thus, a lot in common with how theoretical concepts in the sciences are typically understood and introduced (for more details, see Gebharter 2017b; Schurz and Gebharter 2016). We would also like to emphasize that we do not subscribe to an interventionist interpretation of causation according to which causation is characterized or even defined (see, e.g., Woodward 2003) in terms of interventions.[8] Because interventions are not required for the endeavor of this paper and since how exactly interventions into the kind of mixed systems we are interested in work is still controversial (see, e.g., Baumgartner 2010, 2013; Gebharter 2017a, 2019; Woodward 2015), we prefer an understanding of causation in line with Spirtes et al. (1993) in this paper.

One advantage of a Bayes net representation of causal relations is that the framework comes with a simple test for the productivity of single causally interpreted arrows (cf. Schurz and Gebharter 2016, p. 1087): If $X_i \longrightarrow X_j$ is part of a Bayes net $\langle V, E, P\rangle$, then this arrow is productive if and only if X_j is probabilistically dependent on X_i conditional on X_j's alternative parents $Par(X_j)\backslash\{X_i\}$.[9,10] Now the idea is that a variable D describing an agent's possible decisions has a causal impact on another variable X if and only if $D \longrightarrow X$ is productive, i.e., if there are some circumstances in which D has a probabilistic causal influence on X. In other words: The variable D is a probabilistic causal difference-maker for the variable X. This simple criterion for a variable's causal impact will turn out to be useful when investigating under which conditions the control requirement **(CTRL)** can be satisfied.[11]

[8] For details about the differences between the two kinds of interpretation see, for example, (Glymour 2004).

[9] We define probabilistic dependence of X_j on X_i conditional on a set V' as $P(x_j|x_i, v') \neq P(x_j|v') \wedge P(x_i, v') > 0$ for some x_i, x_j, v'. This is equivalent with the more classical definition which characterizes conditional dependence as $P(x_i, x_j|v') \neq P(x_i|v') \cdot P(x_j|v')$ for some x_i, x_j, v'. Probabilistic independence can be defined as the negation of probabilistic dependence.

[10] Requiring that every arrow in a CBN's graph is productive would amount to assuming minimality (Spirtes et al. 1993, p. 34). For a proof, see (Schurz and Gebharter 2016, Theorem 2).

[11] The same productivity test is used in (Gebharter 2017a) to investigate the possibility of mental causation to the background of causal exclusion worries and in (Gebharter 2019) to explore the reducibility of higher-level and inter-level causation in the presence of mechanistic hierarchies.

5 Free Will and Its Compatibility with the Four Positions

In this section we will translate the four positions introduced in Sect. 3 as well as **(PDO)** and **(CTRL)** into CBN terminology. Hand in hand with the translation process we will develop two causal models which will allow us to evaluate whether **(PDO)** and **(CTRL)** can be combined with these views. We will use these models to explore which combinations of **(PDO)** and **(CTRL)** are metaphysically possible given the different combinations of non-reductive versus reductive physicalism and determinism versus indeterminism. Assuming different combinations of these positions will give us different constraints on the probability distributions compatible with our models and, thus, will delimit the space of metaphysical possibilities. Let us start with some basic assumptions for both models. For the evaluation we are aiming at we need to represent an agent's decisions as well as what is going on at the fundamental physical level. Let us do this by means of a variable D (modeling an agent's decisions) and the three variables P_0, P_1, and P_2 (representing different physical states). We assume that P_0 is a direct cause of P_1, and that P_1 is a direct cause of P_2. Thus, our two models will feature the causal chain $P_0 \longrightarrow P_1 \longrightarrow P_2$. In addition, let us assume that the agential states modeled by D supervene on the physical states represented by P_1.

Next, let us see which constraints we have to accept if non-reductive physicalism were true. Recall that non-reductive physicalism is committed to **(ONI)**, **(SUP)**, and **(CE)**. (**ONI**) demands that D is on an ontologically different level than the physical states represented by P_0, P_1, P_2. Because of this, D should not be identified with one of these variables:

(**ONI**$_{BN}$) D is not identical with one of the variables P_0, P_1, P_2.

(**SUP**) says that an agent's decisions supervene on physical states. For our model this means that D supervenes on P_1, which gives us the following topological and probabilistic constraints:

(**SUP**$_{BN}$) $P_1 \Longrightarrow D$ is part of the model and the following two equations hold:

$$\forall d \forall d' \exists p_1 \; : \; \text{If } d \neq d', \text{ then } P(p_1|d) \neq P(p_1|d') \tag{2}$$

$$\forall p_1 \exists d \; : \; P(d|p_1) = 1 \tag{3}$$

The arrow $P_1 \Longrightarrow D$ represents the supervenience relation in our model. It is required to account for the dependence of D on P_1 and is assumed to technically work exactly like an ordinary single-tailed causal arrow: It conforms to the Markov factorization [Eq. (1)].[12] Equation (2) says that if the decision variable D changes

[12] For an argument to back up this claim, see (Gebharter 2017a). Note that though supervenience has much in common with causation, it is typically considered as some kind of non-causal dependence. One of the main reasons for this is that causes are traditionally assumed to precede their effects in time, while supervening phenomena often occur at the same time as their supervenience bases. Here is another difference: It is typically assumed that a cause's effects can be decoupled from its influence by surgical

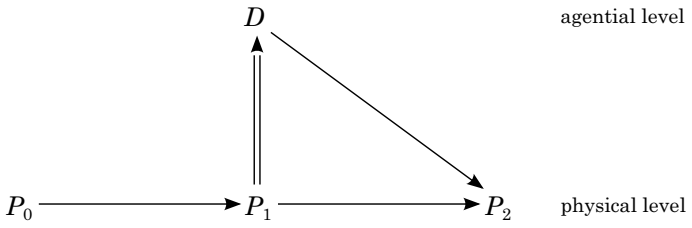

Fig. 1 Graph of the model representing non-reductive physicalism

Fig. 2 Graph of the model representing reductive physicalism

its value, the probability distribution over its supervenience base P_1 has to change as well, and Eq. (3) guarantees that fixing the value of the supervenience base P_1 fully determines the value of the supervening variable D. Note that (**SUP**$_{BN}$) does not provide a definition of supervenience; the condition should rather be seen as a consequence (formulated in terms of the CBN model) that comes with assuming supervenience (**SUP**).

(**CE**) finally forces us to add an arrow from D to P_2. This reflects the assumption that, according to non-reductive physicalism, decisions are at least sometimes causally relevant for an agent's actions at the physical level:

(**CE**$_{BN}$) $D \longrightarrow P_2$ is part of the model.

The graph of the model we get when subscribing to non-reductive physicalism will, according to the considerations above, be the one depicted in Fig. 1.[13] Reductive physicalism, on the other hand, can be modeled by a CBN with the structure depicted in Fig. 2. We arrive at this graph by removing the arrow $P_1 \Longrightarrow D$ and by identifying D with P_1. This move can be justified by translating **(RED)** into (**RED**$_{BN}$). (**RED**$_{BN}$) reflects the main thesis of reductionism which implies that agential properties are identical to physical properties in our model:[14]

Footnote 12 (continued)
interventions. Constraining the probability distributions compatible with our model via (**SUP**$_{BN}$), however, makes it impossible to decouple D from the influence of its supervenience base P_1.

[13] We do not require the arrow $D \longrightarrow P_2$ to be productive at this point because we want to capture non-reductive physicalist positions allowing for inefficacious higher-level causal properties that are still causally relevant via riding piggy back on supervenience and lower-level causation as well.

[14] Note that since D is, according to reductionism, assumed to be identical with P_1, this view will also subscribe to Eqs. (2), (3), and **(CE)** simply because every variable trivially supervenes on itself and P_1 is assumed to be a direct cause of P_2 anyway.

(**RED**$_{BN}$) D is identical with P_1.

We have now two models available: One captures non-reductive physicalism (Fig. 1), and one reductive physicalism (Fig. 2). We can further constrain both models' probability distributions in such a way that they reflect the assumption of determinism or indeterminism. Let us start with determinism. To represent **(DET)** in our models, we have to establish a deterministic dependence of every physical variable on its direct physical cause:

(**DET**$_{BN}$) For all p_i, p_j with $i \in \{0, 1\}$ and $j = i + 1$: $P(p_j|p_i)$ is either 1 or 0.

Indeterminism can, in accordance with **(IND)**, be formulated as the negation of (**DET**$_{BN}$):

(**IND**$_{BN}$) For some p_i, p_j with $i \in \{0, 1\}$ and $j = i + 1$: $P(p_j|p_i)$ is neither 1 nor 0.

One might wonder whether determinism and indeterminism are not represented in a too simplistic way. **(DET)** and **(IND)** both refer to how things are earlier and to the laws of nature. In our models, on the other hand, P_1 and P_2 have only one direct cause and (**DET**$_{BN}$) simply assumes that each P_0-value determines some P_1-value with probability 1 and that each P_1-value determines some P_2-value with probability 1, where (**IND**$_{BN}$) is the negation of that claim. In addition, neither (**DET**$_{BN}$) nor (**IND**$_{BN}$) mention laws of nature. The idea here is the same as in the case of (**SUP**$_{BN}$) above: (**DET**$_{BN}$) and (**IND**$_{BN}$) should not be read as definitions of determinism and indeterminism, respectively, but rather as consequences one gets by assuming **(DET)** or **(IND)** formulated in terms of the two models. The laws of nature, for example, are not explicitly represented in the models. They appear, however, implicitly: They are crucial for how exactly the models' probability distributions are constrained. And that we assume that P_1's only cause is P_0 and that P_2's only cause is P_2 is a harmless simplification. One could also split each one of the variables P_i (with $0 \leq i \leq 2$) into two or more variables. This would make the reasoning and the models a little bit more complicated, but would have no impact on the results we will present below. Hence, we stick with the nice and simple models whose graphs are depicted in Figs. 1 and 2.

Before we can evaluate whether a strong notion of free will committed to **(PDO)** and **(CTRL)** is compatible with the four ontological positions introduced in Sect. 3, we finally have to translate **(PDO)** and **(CTRL)** into the language of our models as well. **(PDO)** requires that the agent has the possibility to decide otherwise. In terms of our two causal models this idea could be expressed by assuming that D's value is not fully determined by the variables describing the past:

(**PDO**$_{BN}$) The conditional probabilities $P(d|x)$ are not extreme, where X stands for the set of variables describing D's past.

Note that (**PDO**$_{BN}$) nicely fits the strong version of the possibility to do otherwise we introduced in Sect. 2: If an agent's decisions are not fully determined by the laws and the past, then the probabilities $P(d|x)$ in (**PDO**$_{BN}$) should not be extreme. This corresponds, in terms of the model, to the assumption that the agent possesses the actualist possibility to do otherwise as is captured by **(PDO)** introduced in Sect. 2.

The second requirement for free will which we introduced in Sect. 2 is that an agent is supposed to have to some extent (causal) control over her environment. To decide whether D has probabilistic causal impact on the agent's environment (here represented by P_2), we can employ the productivity test for single causal arrows introduced in Sect. 4. **(CTRL)** can then be translated as follows:

(**CTRL**$_{BN}$) P_2 probabilistically depends on D conditional on $Par(P_2)\backslash\{D\}$.

More informally, (**CTRL**$_{BN}$) requires that—for an agent to have free will—her decisions must be probabilistic causal difference makers. They have to make a probabilistic causal difference for her environment in at least some causal setting, where possible causal settings are modeled by P_2's alternative causes $Par(P_2)\backslash\{D\}$ taking different values r.

We have now all the tools available for evaluating whether a strong notion of free will subscribing to **(PDO)** and **(CTRL)** is compatible with our four world views introduced in Sect. 3. Let us start with the first model (Fig. 1) representing non-reductive physicalism, which seemed to be one of the more promising views for a strong incompatibilist notion of free will.

Non-reductive Physicalism Plus Determinism According to our first wild guess (Sect. 3) this combination seems, at first glance, to allow for **(PDO)** as well as for **(CTRL)**. However, to get (**PDO**$_{BN}$), D's value must be allowed to vary when conditionalizing on D's past (including P_0). But once we conditionalize on any P_0-value, we get an extreme probability distribution over P_1 due to (**DET**$_{BN}$), which gives us, in turn, an extreme probability distribution over D due to (**SUP**$_{BN}$). Hence, the actualist possibility to do otherwise is excluded if non-reductive physicalism and determinism are true.

Here is an argument that shows that also (**CTRL**$_{BN}$) is excluded if non-reductive physicalism and determinism are true:[15] According to (**DET**$_{BN}$), for every P_1-value p_1 there is a P_2-value p_2 such that $P(p_2|p_1) = 1$ holds, while $P(p'_2|p_1) = 0$ holds for all other P_2 -values p'_2. Now there are two possible cases for every D-value d: Either (i) d and p_1 are compatible (i.e., $P(d, p_1) > 0$), or (ii) they are not (i.e., $P(d, p_1) = 0$). If (i), then $P(p_2|d, p_1) = 1$ and $P(p'_2|d, p_1) = 0$ hold because conditionalizing on compatible values of additional variables does not change conditional probabilities of 1 or 0. Hence, no P_2-value will depend on d given p_1. If (ii), then $P(d, p_1) = 0$. In this case, no P_2-value will depend on d conditional on p_1 by definition. Since d was arbitrarily

[15] Gebharter (2017a) argues that a similar line of reasoning also applies in the context of mental causation and causal exclusion worries. Also note that probabilities could be fine-tuned in such a way that both arrows, $D \longrightarrow P_2$ and $P_1 \longrightarrow P_2$ turn out as unproductive. For why this result should still be counted against the efficacy of the higher-level variable, see (ibid).

chosen and also p_1 was arbitrarily chosen, it follows that P_2 is probabilistically independent of D conditional on $Par(P_2)\backslash\{D\} = \{P_1\}$. But this just means that the arrow $D \longrightarrow P_2$ is not productive, i.e., that D cannot have any causal impact on P_2 whatsoever, which contradicts (**CTRL**$_{BN}$).

Non-reductive Physicalism Plus Indeterminism Does non-reductive physicalism combined with indeterminism fare better when it comes to the question of free will? Since neither the agential nor the fundamental physical level are assumed to be deterministic, an agent could be able to do otherwise in this view. According to (**PDO**$_{BN}$), everything required for this is that D is not fully determined by its past. This actually turns out to be true in our model. Even if one conditionalizes on P_0, nothing excludes that the conditional probabilities $P(d|p_0)$ are non-extreme. If the probability distribution of our model (Fig. 1) would, for example, be specified as follows, (**ONI**$_{BN}$), (**SUP**$_{BN}$), (**CE**$_{BN}$), (**IND**$_{BN}$), and (**PDO**$_{BN}$) would be satisfied and $0 < P(d|p_0) < 1$ would hold for arbitrarily chosen D- and P_0-values d and p_0, respectively:

$$
\begin{array}{llll}
P(P_0 = 0) = 0.5 & P(P_1 = 0|P_0 = 0) = 0.75 & P(P_2 = 0|D = 0, P_1 = 0) = 0.75 \\
P(P_0 = 1) = 0.5 & P(P_1 = 1|P_0 = 0) = 0.25 & P(P_2 = 1|D = 0, P_1 = 0) = 0.25 \\
 & P(P_1 = 0|P_0 = 1) = 0.25 & P(P_2 = 0|D = 0, P_1 = 1) = 0.5 \\
 & P(P_1 = 1|P_0 = 1) = 0.75 & P(P_2 = 1|D = 0, P_1 = 1) = 0.5 \\
 & P(D = 0|P_1 = 0) = 1 & P(P_2 = 0|D = 1, P_1 = 0) = 0.5 \\
 & P(D = 1|P_1 = 0) = 0 & P(P_2 = 1|D = 1, P_1 = 0) = 0.5 \\
 & P(D = 0|P_1 = 1) = 0 & P(P_2 = 0|D = 1, P_1 = 1) = 0.25 \\
 & P(D = 1|P_1 = 1) = 1 & P(P_2 = 1|D = 1, P_1 = 1) = 0.75
\end{array}
$$

Unfortunately, there are, again, problems with (**CTRL**$_{BN}$) and D's probabilistic causal impact on P_2: Because of (**SUP**$_{BN}$) for every P_1-value p_1 there is a D-value d such that $P(d|p_1) = 1$ and $P(d'|p_1) = 0$ for all other D-values d'. This time there are two possible cases for every P_2-value p_2: Either (i) p_2 and p_1 are compatible, or (ii) they are not. If (i), then $P(d|p_2, p_1) = P(d|p_1) = 1$ and $P(d'|p_2, p_1) = P(d'|p_1) = 0$. Thus, no D-value depends on p_2 conditional on p_1. If (ii), then no D-value depends on p_2 conditional on p_1 by definition. This result generalizes: Because p_2 and p_1 were arbitrarily chosen, D and P_2 are independent given $Par(P_2)\backslash\{D\} = \{P_1\}$. This means, again, that D cannot have any probabilistic causal impact on P_2 and that one of the requirements for a strong notion of free will we are interested in in this paper is excluded.

Reductive Physicalism Plus Determinism According to this view, D has to be identified with P_1 (see Fig. 2). Nothing we have assumed excludes that P_2 probabilistically depends on D conditional on $Par(P_2)\backslash\{D\}$, which is the empty set if reductionism is true. Hence, (**CTRL**$_{BN}$) can be satisfied and an agent's decisions might have probabilistic causal impact. The following distribution, for example, would allow for (**RED**$_{BN}$), (**DET**$_{BN}$), and (**CTRL**$_{BN}$) to be satisfied:

$$
\begin{array}{lll}
P(P_0 = 0) = 0.5 & P(P_1 = 0|P_0 = 0) = 1 & P(P_2 = 0|P_1 = 0) = 1 \\
P(P_0 = 1) = 0.5 & P(P_1 = 1|P_0 = 0) = 0 & P(P_2 = 1|P_1 = 0) = 0 \\
 & P(P_1 = 0|P_0 = 1) = 0 & P(P_2 = 0|P_1 = 1) = 0 \\
 & P(P_1 = 1|P_0 = 1) = 1 & P(P_2 = 1|P_1 = 1) = 1
\end{array}
$$

The problem for free will if reductionism plus determinism is true is ($\mathbf{PDO}_{BN}$) which requires that D's value is not fully determined by the past, i.e., by P_0. But this is logically excluded by ($\mathbf{DET}_{BN}$). This result does not suit strong incompatibilists or libertarians, but it suits strong compatibilists or soft determinists.

Reductive Physicalism Plus Indeterminism Here comes the last one of the possible combinations to be explored. This view is the only one of the four combinations of positions discussed in this paper that actually can account for libertarian free will committed to the possibility to do otherwise as well as the probabilistic causal efficacy of an agent's wilful decisions. ($\mathbf{PDO}_{BN}$) can be satisfied because ($\mathbf{IND}_{BN}$) allows for non-extreme conditional probabilities $P(p_1|p_2)$, and ($\mathbf{CTRL}_{BN}$) can be satisfied simply because nothing excludes a dependence of P_2 on P_1 in the second model (see Fig. 2). This can be demonstrated by the following probability distribution which allows for ($\mathbf{RED}_{BN}$), ($\mathbf{IND}_{BN}$), and ($\mathbf{PDO}_{BN}$) as well as for ($\mathbf{CTRL}_{BN}$) to be satisfied:

$$\begin{array}{rrrr} P(P_0=0)=0.5 & P(P_1=0|P_0=0)=0.75 & P(P_2=0|P_1=0)=0.75 \\ P(P_0=1)=0.5 & P(P_1=1|P_0=0)=0.25 & P(P_2=1|P_1=0)=0.25 \\ & P(P_1=0|P_0=1)=0.25 & P(P_2=0|P_1=1)=0.25 \\ & P(P_1=1|P_0=1)=0.75 & P(P_2=1|P_1=1)=1.75 \end{array}$$

Summarizing, it turned out that three of the four possible combinations of the positions discussed exclude a notion of free will committed to both **(PDO)** and **(CTRL)**. Both versions of non-reductive physicalism, the deterministic as well as the indeterministic one, exclude an agent's ability to control her environment **(CTRL)**. **(PDO)**, on the other hand, is only excluded in the deterministic setting. So in the end, given non-reductive physicalism were true, only indeterminism does not exclude one of the requirements for free will discussed, viz. **(PDO)**. If reductionism were true, on the other hand, **(CTRL)** is always satisfiable, while **(PDO)** is only satisfiable in the indeterministic setting and excluded in the deterministic setting. The following table summarizes these results:

	Determinism	Indeterminism
Non-reductive physicalism	×	**(PDO)**
Reductive physicalism	**(CTRL)**	**(PDO) + (CTRL)**

6 Conclusion

This paper was intended as a first step into the bigger project of investigating what can be learnt about free will from the point of view of an empirically informed understanding of causation such as the CBN framework. We think that it makes sense to start such an endeavor with strong incompatibilist notions of freedom, committed to an actualist interpretation of the ability to do otherwise **(PDO)** as well as to the view that an agent can control her environment via her willful decisions exerting probabilistic causal influence **(CTRL)**. In particular, we

investigated whether this kind of freedom is compatible with non-reductive and reductive physicalism plus determinism or indeterminism. We then constructed two causal models on whose basis we have shown that three of the four possible combinations of these positions lead to problems with such strong incompatibilist notions of freedom. It turned out that only reductionism together with indeterminism allows for both **(PDO)** and **(CTRL)**. According to our analysis, liberterians subscribing to both are safe for the moment. If one is, however, not ready to subscribe to reductionism and indeterminism, one needs to subscribe to a weaker notion of free will. If one is satisfied with a notion of freedom that only requires **(PDO)**, then one might want to subscribe to indeterminism, and if one prefers **(CTRL)** over **(PDO)**, subscribing to reductionism seems to be the way to go.

We consider the result that non-reductive physicalism is not as attractive as one might think at first glance as quite surprising. (Compare also the two tables in Sects. 3 and 5.) While supporters of non-reductive physicalism have only one of the conditions of strong accounts of free will, **(PDO)**, available, and even this only if they are ready to subscribe to indeterminism as well, reductionists can get free will (in the sense of **(CTRL)** alone) as well as the strongest one of the notions investigated ((**PDO)** plus **(CTRL)**). They can get the former by subscribing to determinism, and the latter by going for indeterminism. One can use these results for building different arguments. One could, for example, use it to persuade anyone flirting with determinism to be reductionist, or to push libertarians to subscribe to reductive physicalism rather than to non-reductive physicalism. Vice versa, one could see the result as support for one of the four ontological positions depending on which notion of free will one presupposes or finds evidence for.

There might, of course, be other problems with freedom if reductionism and/or indeterminism were true such as the prominent worry that indeterminism would render the effects of an agent's decisions to some extent a matter of luck (see, e.g., Strawson 1994; Kane 1996; Mele 2006). We conjecture that modern theories of probabilistic actual causation (see, e.g., Fenton-Glynn 2017) might be able to shed new light on the luck problem. Such approaches allow for actual causation in an indeterministic world and, thus, probably for the formulation of degrees of causal influence even if an agent's decisions do not fully determine her actions. Issues like this as well as questions concerning the compatibility of weaker versions of freedom with different views about the world from a causal modeler's perspective have to await their investigation in future work.

Acknowledgements This work was supported by Deutsche Forschungsgemeinschaft (DFG), research unit Inductive Metaphysics (FOR 2495). We would like to thank Randolph Clarke, Nadine Elzein, Christian Feldbacher-Escamilla, Christian Loew, and Kadri Vihvelin for important discussions and Seán Levey for helpful comments on an earlier version of this paper. Thanks also for valuable comments from the audiences at the PCCP meeting at the Department of Theoretical Philosophy, University of Groningen, and at the DCLPS research colloquium, University of Düsseldorf. Finally, we would like to thank two anonymous reviewers for helpful comments.

References

Austin, J. L. (1979). Ifs and cans. In J. O. Urmson & W. G. J (Eds.), *Philosophical papers* (pp. 205–232). Oxford: Oxford University Press.

Ayer, A. J. (1954). Freedom and necessity. *Philosophical essays* (pp. 271–284). London: Macmillan.

Baumgartner, M. (2010). Interventionism and epiphenomenalism. *Canadian Journal of Philosophy*, *40*(3), 359–383.

Baumgartner, M. (2013). Rendering interventionism and non-reductive physicalism compatible. *Dialectica*, *67*(1), 1–27.

Chisholm, R. (1964/2015). Human freedom and the self. In J. Dancy & C. Sandis (Eds.), *Philosophy of action: An anthology* (pp. 347–352). Oxford: Blackwell.

Chisholm, R. (1976). The agent as cause. In M. Brand & D. Walton (Eds.), *Action theory* (pp. 199–211). Dordrecht: Reidel.

Clarke, R. (2003). *Libertarian accounts of free will*. Oxford: Oxford University Press.

Dennett, D. C. (1984). I could not have done otherwise—So what? *Journal of Philosophy*, *81*(10), 553–565.

Ekstrom, L. (1998). Protecting incompatibilist freedom. *American Philosophical Quarterly*, *35*(3), 281–291.

Ekstrom, L. (2003). Free will, chance, and mystery. *Philosophical Studies*, *113*(2), 153–180.

Elzein, N., & Pernu, T. (2017). Supervenient freedom and the free will deadlock. *Disputatio*, *9*(45), 219–243.

Eronen, M. I., & Brooks, D. (2018). Levels of organisation in biology. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/spr2018/entries/levels-org-biology/.

Fara, M. (2008). Masked abilities and compatibilism. *Mind*, *117*, 843–865.

Fenton-Glynn, L. (2017). A proposed probabilistic extension of the Halpern and Pearl definition of 'actual cause'. *British Journal for the Philosophy of Science*, *68*(4), 1061–1124.

Fischer, J. M., & Ravizza, M. S. J. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge: Cambridge University Press.

Frankfurt, H. (1969). Alternate possibilities and moral responsibility. *Journal of Philosophy*, *66*, 829–839.

Frankfurt, H. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy*, *68*, 5–20.

Gebharter, A. (2017a). Causal exclusion and causal Bayes nets. *Philosophy and Phenomenological Research*, *95*(2), 153–375.

Gebharter, A. (2017b). *Causal nets, interventionism, and mechanisms*. Cham: Springer.

Gebharter, A. (2017c). Uncovering constitutive relevance relations in mechanisms. *Philosophical Studies*, *174*(11), 2645–2666.

Gebharter, A. (2019). A causal Bayes net analysis of Glennan's mechanistic account of higher-level causation (and some consequences). *British Journal for the Philosophy of Science*. https://doi.org/10.1093/bjps/axz034.

Gebharter, A. (ms). *Freedom as a higher-level phenomenon*?

Ginet, C. (1990). *On action*. Cambridge: Cambridge University Press.

Glymour, C. (2004). Critical notice. *British Journal for the Philosophy of Science*, *55*(4), 779–790.

Hausmann, M. (2019). The consequence of the consequence argument. *Kriterion - Journal of Philosophy*. Retrieved from http://www.kriterion-journal-of-philosophy.org/kriterion/issues/Permanent/Kriterion-hausmann-01.pdf.

Hitchcock, C. (2012). Theories of causation and the causal exclusion argument. *Journal of Consciousness Studies*, *19*(5–6), 40–56.
Hobart, R. E. (1934). Free will as involving determination and inconceivable without it. *Mind, XLIII*, *169*, 1–27.
Hoefer, C. (2016). Causal determinism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/spr2016/entries/determinism-causal/.
Hume, D. (1748/1999). An enquiry concerning human understanding. Oxford: Oxford University Press.
Kane, R. (1994). Free will: The elusive ideal. *Philosophical Studies*, *75*(1–2), 25–60.
Kane, R. (1996). *The significance of free will*. Oxford: Oxford University Press.
Kane, R. (2007). Libertarianism. In J. M. Fischer (Ed.), *Four views on free will* (pp. 5–43). Oxford: Blackwell.
Kim, J. (2005). *Physicalism, or something near enough*. Princeton: Princeton University Press.
Lewis, D. (1997). Finkish dispositions. *Philosophical Quarterly*, *47*(187), 143–158.
List, C. (2014). Free will, determinism, and the possibility of doing otherwise. *Noûs*, *48*(1), 156–178.
Mele, A. (Ed.). (2006). *Free will and luck*. Oxford: Oxford University Press.
Moore, G. E. (1912). *Ethics*. Oxford: Oxford University Press.
Nahmias, E. (2011). Intuitions about free will, determinism, and bypassing. In R. Kane (Ed.), *The Oxford handbook of free will* (pp. 555–576). Oxford: Oxford University Press.
Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2005). Surveying freedom: Folk intuitions about free will and moral responsibility. *Philosophical Psychology*, *18*(5), 561–584.
Nichols, S. (2011). Experimental philosophy and the problem of free will. *Science*, *331*(6023), 1401–1403.
O'Connor, T. (2000). *Persons and causes: The metaphysics of free will*. New York: Oxford University Press.
Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA: Morgan Kaufmann.
Pereboom, D. (2001). *Living without free will*. Cambridge: Cambridge University Press.
Pink, T. (2016). *Self-determination: The ethics of action*. Oxford: Oxford University Press.
Reichenbach, H. (1956). *The direction of time*. Berkeley: University of California Press.
Sarkissian, H., Chatterjee, A., de Brigard, F., Knobe, J., Nichols, S., & Sirker, S. (2010). Is belief in free will a cultural universal? *Mind and Language*, *25*(3), 346–358.
Saunders, J. (1968). The temptations of 'powerlessness'. *American Philosophical Quarterly*, *5*(2), 100–108.
Schaffer, J. (2016). Grounding in the image of causation. *Philosophical Studies*, *173*(1), 49–100.
Schurz, G., & Gebharter, A. (2016). Causality as a theoretical concept: Explanatory warrant and empirical content of the theory of causal nets. *Synthese*, *193*(4), 1073–1103.
Shabo, S. (2011). Why free will remains a mystery. *Pacific Philosophical Quarterly*, *92*(1), 105–125.
Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search* (1st ed.). Dordrecht: Springer.
Strawson, G. (1986). *Freedom and belief*. Oxford: Oxford University Press.
Strawson, G. (1994). The impossibility of moral responsibility. *Philosophical Studies*, *75*, 5–24.
van Inwagen, P. (1975). The incompatibility of free will and determinism. *Philosophical Studies*, *27*, 185–199.
van Inwagen, P. (1983). *An essay on free will*. Oxford: Oxford University Press.
van Inwagen, P. (2000). Free will remains a mystery. *Philosophical Perspectives*, *14*(1), 1–20.
van Inwagen, P. (2008). How to think about the problem of free will. *Journal of Ethics*, *12*(3–4), 327–341.
Vihvelin, K. (2004). Free will demystified: A dispositional account. *Philosophical Topics*, *32*(1), 427–450.
Watson, G. (1975). Free agency. *Journal of Philosophy*, *72*, 205–220.
Woodward, J. (2003). *Making things happen*. Oxford: Oxford University Press.
Woodward, J. (2015). Interventionism and causal exclusion. *Philosophy and Phenomenological Research*, *91*(2), 303–347.