Alternative Possibilities and the Meaning of 'Can'

Maria Sekatskaya & Gerhard Schurz

Alternative Possibilities and the Meaning of 'Can'

Maria Sekatskaya & Gerhard Schurz

Our account of free will integrates a counterfactual conditional analysis of abilities with a Frankfurt-style sourcehood psychological approach and is compatible with both determinism and indeterminism. It effectively addresses criticisms of the conditional analysis of "can" by demanding coherence between agents' free actions and their personality frames. The paper begins by discussing conditional analyses of abilities, followed by an exploration of three strategies to counter the consequence argument: first, by assuming determinism with a backtracking analysis of counterfactuals; second, by assuming determinism with a local miracle analysis of counterfactuals; and third, by assuming indeterminism. We further demonstrate that the first two strategies we propose are immune to the criticisms faced by other conditional accounts. Moreover, we show that the third strategy effectively solves the luck problem. The paper concludes by affirming the reality of free will and its consistency with a naturalistic worldview.

1 Alternative Possibilities and Conditional Analysis of Abilities

There is a wide agreement in the free will debate that having free will implies possessing the capacity to choose one's course of action. The natural reading of "choosing" seems to demand that an agent choose between alternative possibilities. The contested question, however, is how to interpret these alternative possibilities. Are there alternative possibilities in a deterministic world? Incompatibilists argue that determinism precludes alternative possibilities, and is, therefore, incompatible with free will. This reasoning can be shortly summarized as follows. An agent's act is free only if it is in the agent's power (up to the agent) to choose to act in one way or another, and to act in this way:

(1) *x* acted freely only if *x* could have done otherwise.

Henceforth, we abbreviate the thesis "*x* could have done otherwise" as CDO. CDO implies that agent *x* has alternative possibilities of the right sort (at some time not later than the time of the agent's action) so that he can choose and perform alternative possible actions. Can CDO be true in a deterministic world? According to incompatibilists,

(2) Physical determinism rules out any alternative possibilities

because determinism (D) is defined as the thesis that "there is at any instant exactly one physically possible future" (Inwagen 1983, 3). Hence, if D is true, CDO is false: in a deterministic world no one acts freely.

Compatibilists can reject either thesis (1) or thesis (2). Some compatibilists have rejected (1) (cf. Dennett 1984; Frankfurt 1969). However, this move is rather radical, because denying our capacity to choose otherwise looks suspiciously close to denying free will outright. The classical compatibilist move is to reject (2), which can be done by reading CDO as a conditional statement: "*x* could have done otherwise" means "*x* would have done otherwise if a certain condition *C* obtained".

The conditional analysis of freedom of will was initially proposed by David Hume (1748), and later developed by G.E. Moore (1912), Dickinson Miller (Miller 1934), and Alfred Ayer (1954). It enjoyed wide acceptance among naturalistically inclined analytic philosophers until John Austin's (1961) and Keith Lehrer's (1968) criticisms showed that the versions of the conditional analysis that had been provided so far were flawed. However, if one wants to demonstrate that incompatibilism is right, it is not enough to show that some versions of conditional analysis are wrong. Although thesis (2) might seem intuitively true, if some version of conditional analysis succeeds, (2) will turn out false. Incompatibilists must show that there are strong reasons to believe that physical determinism and alternative possibilities are incompatible. This is the aim of the so-called "consequence argument" (CA), first published by Carl Ginet (1966) and Peter Inwagen (1975).

Before criticizing the CA, which we do in sections 2 and 3 of our paper, in this first section we give a brief review of the recent theories of a classical compatibilist style. These theories propose a conditional analysis of CDO along the following lines: an agent could have done otherwise if he had an ability such that, if condition *C* obtained, and he tried to use this ability, he would have succeeded. After that, we will clarify the notion of ability we rely on. In section 3 we propose three ways to reject the CA: by assuming (i)

determinism with backtracking analysis of counterfactuals, (ii) determinism with local miracle analysis of counterfactuals, and (iii) indeterminism. In section 4 we present our backtracking compatibilist analysis of abilities to do otherwise. In sections 5 and 6 we explain in more detail why our backtracking compatibilist account does not have the problems that some other conditional accounts have. In the rest of the paper, we present our local miracle and our indeterministic compatibilist analyses of abilities to do otherwise, and show that they effectively solve the randomness objection and the luck problem.

"New dispositionalist" compatibilists explain agents' abilities in terms of dispositions to give a certain response to the stimulus of their own trying (Vihvelin 2004, 2013; Fara 2008). While we do agree that agents' abilities can be analyzed in terms of dispositions to give certain responses to particular stimuli that are partly constituted by some relevant psychological state of the agent, we don't assert that it is necessarily a stimulus of the agent's own trying since it has been shown that in some cases the analysis in terms of trying is problematic (Franklin 2011; Kittle 2015b). In our explanation of abilities, we will follow David Lewis (1997), who connected dispositions to give responses to certain stimuli with intrinsic properties of the bearers of these dispositions. In order to avoid the problem with Finkish dispositions or Finkish lack of dispositions, Lewis introduced a time interval during which the intrinsic properties of the bearers of these dispositions should not change. Although Lewis himself did not explicitly use his analysis of dispositions to explain agents' abilities, it can, in our opinion, quite naturally be extended in this way. We assert that an agent's abilities are a specific class of the agent's dispositions to act in particular ways in particular circumstances, where these acts are partly caused by the agent's intrinsic psychological and physical properties, such as the agent's skills, beliefs, desires, etc. Thus, we propose the following definition of having an ability at a time:

> ABILITY. An agent x has at time t the ability to do A iff

(a) x has an intrinsic property B between t and some later time point t', and

(b) if certain conditions C_i ($i = 1, 2, ...$) would obtain between t and later times $t_i \leq t'$, then C_i and x's having of B would jointly be an x-complete

cause of x's doing A (where "an x-complete cause" is "a cause complete in so far as properties intrinsic to x are concerned").[1,2]

What the conditions C_i and property B are depends on the ability in question. If we consider the ability to play the violin, then the conditions C_i consist of proper external circumstances (e.g., having a violin at hand, etc.) plus a proper internal psychological stimulus on the part of the agent, for example, his decision to play or his desire to play the violin; note that different conditions have to endure for different time spans after time t. Property B, on the other hand, consists of the agent's skills, beliefs, etc. Our account will show how, given any suitable understanding of abilities along the lines above, one can explain the abilities to do otherwise, within the framework of either physical determinism or indeterminism. We will use this definition of ability in our own account of CDO in sections 4–8 of this paper, where we will explain what the conditions C_i are for the abilities to do otherwise, by using the framework of possible world analysis of counterfactual conditionals. There it will be clear that our account offers the kind of ability to do otherwise that many theories of free will are after, the one called "all-in ability" (Austin 1961), "wide ability" (Vihvelin 2013), "ability with an opportunity" (Franklin 2011), or "maximally specific ability" (Kittle 2015b).

Abilities, understood in a new dispositionalist way, are compatible with determinism. However, if the CA is sound, then physical determinism implies that no one could have ever done otherwise, and therefore, in a deterministic world no one has the ability to do otherwise. In sections 3–8 we will show how compatibilists can secure their position against the destructive effect of the CA without being vulnerable to standard objections against compatibilism. We will do so by combining conditional analysis with a suitable version of a

---

1 This definition is based on a modified version of Lewis' definition in Lewis (1997, 157); the characterization of an "x-complete cause" is found in Lewis (1997, 156). The main difference between our account of abilities and Lewis' account of dispositions is that Lewis is interested in dispositions of any kind of entities to respond to relevant stimuli. Neither the disposition nor the stimulus must have something to do with agency or the psychological circumstances of the act, which are essential for questions about free will. In our account, the condition C_i has to include the agent's first-order desires, and the intrinsic property B has to include the agent's personality frame, as will be shown in section 4.

2 Further problems of Lewis' (1997) account of dispositions can be fixed but cannot be discussed here. For example, in order to admit a probabilistic or gradual notion of disposition one could follow Vihvelin's proposal (2013, 187) and weaken *Ability* so that the condition following "then" must hold only in a suitable proportion of cases.

sourcehood account (drawing on Frankfurt's (1971) notion of second-order desires as well as Fischer and Ravizza's (1998) condition of reason-reactivity).

2 The Review of the Consequence Argument

The "third" version of the CA, published in van Inwagen's (1983) book, has attracted the most attention in the free will debate. It contains three propositions:

P_0. A proposition that describes the total state of the world at some moment in the distant past (t_0).

L. A proposition that is the conjunction of all the laws of nature.

P. A true proposition about time t_1 after time t_0.

N, a sentential modal operator defined:

Np. p, and no one has, or ever had, any choice about whether p.

Two modal principles, or rules of inference:

RULE ALPHA. If p is a necessary truth, then p is true and no one has, or ever had, any choice about p. ($\Box p \vdash Np$)

RULE BETA. If p and no one has or had any choice about p, and if $p \supset q$ and no one has or had any choice about $p \supset q$, then q and no one has or had any choice about q. ($Np, N(p \supset q) \vdash Nq$).

Using these notations, the argument has the following logical structure:

1.	$\Box((P_0 \wedge L) \supset P)$	Symbolic definition of Determinism
2.	NP_0	Principle of the Fixity of the Past
3.	NL	Principle of the Fixity of the Laws
∴	NP	Conclusion, contradicts CDO

The proof:

4.	$\square(P_0 \supset (L \supset P))$	1, Exp within $\square$-scope
5.	$N(P_0 \supset (L \supset P))$	4, Rule Alpha
6.	$N(L \supset P)$	2, 5, Rule Beta
$\therefore$	NP	3, 6, Rule Beta

Since P can be any true proposition about what someone does, NP asserts that no one has any choice about any of her actions. If a compatibilist wants to reject the CA, she has to reject either one of the inference principles or one of the premises.

Rule Alpha is very plausible and has been widely accepted in this discussion.

Rule Beta is known as the Principle of Transfer of Powerlessness and, according to the proponents of the CA, is also very plausible (cf. Ginet 1980, 182; Inwagen 1983, 99). However, Rule Beta is the most disputed part of the CA, usually criticized by means of counterexamples (Widerker 1987; McKay and Johnson 1996; Carlson 2003). McKay and Johnson argue that Alpha and Beta together entail the Principle of Agglomeration: $Np, Nq \vdash N(p \wedge q)$.[3] However, this principle can be shown as invalid by applying the condition "$N(\text{-})$", that no one can do anything about, to the outcome of a random process, viz. the tossing of a fair coin. After criticizing van Inwagen's formulation of Beta they propose four different modal principles closely resembling Beta, which are immune to this counterexample but still can be used in deriving the conclusion of the CA. It has been argued that these different principles are less intuitive than the original Beta and have unwelcome consequences (Blum 2003). Other versions of Beta have been proposed (Carlson 2000, 2003; Crisp and Warfield 2000) and currently the discussion is very much alive (Gustafsson 2017).

Van Inwagen himself reacted to McKay and Johnson's (1996) counterexample by conceding that his version of Beta was invalid, and by modifying the N-operator as follows:

---

3 McKay and Johnson give the proof (1996, 115):

1. Np (premise)
2. Nq (premise)
3. $\square[p \supset (q \supset (p \wedge q))]$ (necessity of a logical truth)
4. N$[p \supset (q \supset (p \wedge q))]$ (from 3 and α)
5. N$[q \supset (p \wedge q)]$ (from 1, 4 and β)
6. N$(p \wedge q)$ (from 2, 5 and β)

> *p* and every region to which anyone has, or ever had, *exact* access
> *is a subregion of p*. One has *exact* access to a region if one has
> access to it *and to none of its proper subregions*. Intuitively, one has
> exact access to *p* if one can ensure the truth of *p* but of nothing
> "more definite". (Inwagen 2000, 8)

So, according to this definition, McKay and Johnson's (1996) case is not a counterexample to Rule Beta anymore. An agent can have exact access to the region of logical space in which "The coin is tossed and it lands either heads or tails" holds, but not to its subregions where specifically "The coin is tossed and it lands heads" holds, or where specifically "The coin is tossed and it lands tails" holds.

Lynn Baker has argued that the new *N*-operator leads to the conclusion that "every region of logical space to which anyone has, or ever had, exact access is the region containing only the actual world" (2008, 16). If this is correct, then the conclusion of the CA will follow quite independently from the assumption of determinism. However, we will not explore the implications of the new *N*-operator, because in the following sections we will show that compatibilists can deny NP_0 and *NL* on both readings of *N*.

Moreover, we claim that supporters of agents' abilities to do otherwise do not need to withdraw (suitable versions of) Rule Beta if a version of Beta that withstands objections can be formulated. Rather, they can and should reject one of the three premises:

- If determinism is accepted, then either 2. or 3. is to be rejected, as in classical compatibilist and new dispositionalist positions,
- If indeterminism is accepted, then 1. must be rejected, as in libertarian positions.

In the next sections, we show how these premises can be rejected by developing a new kind of conditional analysis of freedom using the formal tool of the analysis of counterfactual conditional statements in terms of possible worlds. We argue that our proposal has three advantages compared to traditional accounts:

1. It answers the standard objection against conditional explications of CDO that they are too weak to remove counterexamples, by combining them with a suitable version of a sourcehood account based on the

condition of coherence with one's personality frame, abbreviated as (CPF).

2. It is flexible enough to be compatible with two versions of determinism, (i) a backtracking variant and (ii) a local-miracle variant, as well as with (iii) indeterminism.

3. Moreover, the condition CPF may also solve specific problems of the three versions, e.g., the "anything possible" objection against (ii) and the problem of luck objections against (ii) and (iii).

Some compatibilists have provided arguments against the CA (cf. Vihvelin 1988; Taylor and Dennett 2002) and developed their own compatibilist theories of free will (Vihvelin 2013; Dennett 2003). However, an analysis of the precise connection between a refutation of the CA and a formulation of a conditional compatibilist analysis of CDO is missing so far. In this paper, we are going to address this issue. We will incorporate the results of the previous critics of the CA into one unified framework that is based on a counterfactual analysis and shows how both a backtracking and a local miracle analysis of counterfactuals can be used to refute the CA and to provide a positive account of CDO. Moreover, we will show that the same counterfactual framework can be used to explicate an indeterministic account of CDO that drops the assumption of determinism instead of employing backtracking or local miracles.

We chose the version of the CA that we did because it is arguably one of the strongest arguments against compatibilism (cf. Capes 2019). Our proposed refutation of this version of the CA also works against the "Basic Version" of the CA. The Basic Version depends on the acceptance of the "Extension Principle": "An agent can do X only if his doing X can be an extension of the actual past, holding the laws fixed" (Fischer 1994, 88). The "Extension Principle" is a straightforward affirmation of the Fixity of the Past and the Fixity of the Laws, and in sections 3–7 we show how both Fixity Principles can be consistently denied.

3 Compatibilist Rejection of the Fixity Principles

In the following sections 3–7 we assume determinism and propose our new explication of compatibilist conditional freedom within a deterministic framework. Premise 2., the Principle of the Fixity of the Past, states that P_0, a proposition that describes the total state of the world at some moment in the

distant past, is true and no one can make P_0 false where "x makes P_0 false" is understood in the following weak sense: if x had acted otherwise, then the distant past would have been different.[4] This formulation does not assume that there is a direct causal relationship between the agent's actions and the change of the events in the remote past.

John Saunders (1968) was the first to reject the Fixity of the Past with this kind of strategy, which later became known as backtracking (Fischer 1988). In our viewpoint, the main advantage of the backtracking strategy in the context of the freedom debate is that it applies to the conditional analysis of "can". According to the backtracking strategy, "P_0 and no one has, or ever had, any choice about whether P_0" is wrong, because the agent, x, can perform not-P now, and if x performs not-P now then it would have been false that P_0.[5] So x has the power to change the past P_0 in the weak sense explained above. The connection between the backtracking strategy and the conditional analysis of "can" is that the causal chain leading from the counterfactual alteration of P_0 to the counterfactual alteration of P involves a counterfactual alteration of the agent's will (decision) at some intermediate time.

The backtracking strategy as applied to the abilities of agents is just a particular case of the backtracking analysis of the truth of counterfactual statements of the form $(P > Q)$. According to Jonathan Bennett, "$(P > Q)$ is true iff Q is true at all the P-worlds which are closest to the actual world" (1984, 57), and since we want, even in deterministic worlds, some counterfactuals to be true and some false, we have two options to choose from:

> [...] if P is false (at the actual world), then every causally possible P-world is unlike the actual world in respect of its whole history up to the time (T) to which P pertains. Any good statement of the determinist thesis will tell you that much, making it clear that any two worlds which are strictly determined by the same laws are unlike at time T only if they are unlike at every earlier time. So, if we want to evaluate $(P > Q)$ where P is false, we must either

---

4 On van Inwagen's new formulation of N, this should be read as "P_0 and every region to which anyone has, or ever had, exact access is a sub-region of P_0". The rest of our argument applies equally well to the old and the new formulation of the N-operator. The difference is that where we say "x has a choice about whether P_0" on the old formulation of N, we substitute "x has exact access to a region where P_0 is false" on the new formulation of N.

5 By "performing P" we mean performing an act such that "P" is a proposition describing this act, and by "performing not-P" we mean performing an act such that "not-P" is a proposition describing this act.

accept as "closest" some worlds which are unlike ours at all times
earlier than T, or deem to be "closest" some worlds which are
just like ours up to about T and are then pushed off our course by
a miracle—an event breaking some actual causal law. (Bennett
1984, 59)

Bennett chooses the first option, Lewis (1979) chooses the second. The back-
tracking analysis rejects Premise 2., i.e., assumes a different past globally, i.e.,
in many instances. In contrast, Lewis' local miracle strategy rejects Premise
3., i.e., requires a violation of a law but only locally, i.e., in only one instance.

In section 4 we will elaborate on how the backtracking analysis of counter-
factuals together with the conditional analysis of abilities yields a compatibilist
analysis of free will. In a nutshell, the idea is the following: we say that x
could have done otherwise than P at t_n if there are possible worlds close to
the actual world at t_{n-1} in some respect to be clarified in what follows, such
that in these worlds x does otherwise than P at t_n.

The local miracle account is also a possible way to go for a compatibilist,
viz. to reject Premise 3., the Principle of the Fixity of the Laws. It states that no
one can change the laws of nature (has a choice about what the laws of nature
are), where "x changes the laws of nature" is understood in the following
weak sense: if x had acted otherwise then the laws of nature would have been
different. This formulation is a slight reformulation of Lewis' (1981) weak
compatibilist thesis in a way that avoids van Inwagen's (2004) criticism.

David Lewis (1981) distinguished between two senses in which a law of
nature can be broken in connection with what an agent does. In a strong
sense, it can be broken by an action that an agent performs or by a direct
consequence of an action that an agent performs. For example, a law is broken
in a strong sense if an agent moves his hand faster than the speed of light
or throws a stone that flies faster than the speed of light. Crediting an agent
with this kind of ability is implausible, so, read in the strong sense, Premise
3., NL, is true. However, a law of nature can be broken in a weak sense: it is
possible that somewhere in the past a local miracle happened. In this case, it
is possible that an agent does otherwise than P as a consequence of this prior
miracle, which could have happened at any time between P_0 and P.

Lewis made his objection against an earlier version of the CA (Inwagen
1975) in the following way:

(Weak Thesis) I am able to do something such that, if I did it, a
law would be broken.

(Strong Thesis) I am able to break a law. (Lewis 1981, 115)

The Weak Thesis, which as a soft determinist I accept, is the thesis that I could have rendered a law false in the weak sense. The Strong Thesis, which I reject, is the thesis that I could have rendered a law false in the strong sense. (Lewis 1981, 120)

According to Lewis, it is the strong sense that is incredible, and it is the weak sense that follows from the CA, so the CA is not a problem for a compatibilist (soft determinist) position. Van Inwagen objected that even the Weak Thesis is incredible because it ascribes to an agent the power to perform miracles, where a miracle is defined as "an event or state of affairs whose occurrence would be inconsistent with the whole truth about the past and the laws of nature" (Inwagen 2004, 349). But this incredibility is arguably due to the inappropriateness of the phrase "a law of nature is broken". We think that what Lewis means with this is nothing more than what we said above, namely, that the laws of nature are different in the actual world (where I perform P) and the counterfactual world where I act differently. Tognazzini (2016) argues in more detail why this is what Lewis must have meant with a law being "broken", and what "miracles" are according to Lewis. Helen Beebee and Alfred Mele (2002) argue that Humeanism about laws of nature supports not only the Weak, but also the Strong Thesis, and this is a problem for Lewis' local miracle compatibilism. However, in our local miracle compatibilist proposal, presented in section 7, we do not endorse Humeanism about laws of nature. Therefore we are free from the problems discussed in Beebee and Mele (2002). In section 7 we also show how our account solves the problems discussed in Beebee (2003).

Both the backtracking strategy and the local miracle strategy are legitimate ways for a compatibilist to reject the CA. Indeed, the compatibilist accepts *ex hypothesi* that $(P_0 \wedge L) \supset P$ and that P_0, L, and P are true in the actual world. The decisive question is: what is involved in making the conditions C_i true in a possible world sufficiently close to the actual world? The compatibilist who accepts determinism can give two answers according to our analysis, one based on the backtracking strategy and the other one on the local miracle strategy. The indeterminist, in contrast, can explicate the conditions C_i in a way that neither implies a global change of the past nor a local miracle. In the next five sections, we elaborate on these three options.

4 Backtracking Compatibilist Proposal

If the actual world a is deterministic, the backtracking strategy allows us to formulate the following conditions that have to obtain for CDO to be true about a person x in a.

> CDO$_B$. x could have done otherwise than P at t_n if x does not-P at t_n in some possible world w that satisfies the following conditions:

a. w and a are governed by deterministic laws that are identical.
b. The pasts of w and a are different at all past times.
c. x's personality frame agrees in a and w at all times until t_{n-1} and it does not change between t_{n-1} and t_n.
d. (1) x's internal state at t_{n-1} in w differs from the corresponding internal state of x in a in regard to some FODs of x, in coherence with x's personality frame, where
   (2) w and a agree in all agent-external facts at t_{n-1} that were causally relevant to x's actual action at t_n.

Characteristics required by CDO$_B$ (a) and (b) were discussed in previous sections. Characteristics CDO$_B$ (c) and (d) need explanation.

Concerning CDO$_B$ (c): According to our account, in order to have free will an agent must have a personality frame (F), which, in turn, includes reasoning and volitional abilities meeting minimal rationality conditions. Minimal rationality conditions demand that an agent meet the criteria of moderate reasons-responsiveness, i.e., have a certain level of reasons-receptivity and reasons-reactivity, as discussed in Fischer and Ravizza (1998). Necessary volitional abilities include an agent's capacity to form first- and second-order desires and volitions. Following Harry Frankfurt, we define a first-order desire (FOD) as a desire "to do or not to do one thing or another" (1971, 7), a second-order desire (SOD) as a desire that a certain FOD become causally efficient (1971, 10), a first-order volition as an effective FOD that causally contributes to an agent's act (1971, 8), and a second-order volition as a SOD that is a part of the cause of the agent's first-order volition (1971, 10). A person having abilities at a time should be understood as explained in Ability. A person x's having the ability to do A at t_n is a necessary condition for the truth of the claim that "x could have done A at t_n". For example, if we want to know whether the claim "x could have played the piano at t_n instead of playing the violin" is true, we should consider the possible worlds where x has the same

intrinsic property B (causal basis of x's ability to play the piano) and where conditions C_i vary without violating CDO_B (d).

In addition to the necessary abilities listed above, common to all persons with free will, each F has a particular set of characteristics, including this person's SODs, essential FODs such as the desire to live, stable character traits, and general and specific abilities and skills, such as an ability to play the violin, an ability to play the violin in front of a big audience, an ability to play the violin in front of a big audience while being tired, and all the rest of this person's abilities in all the ranges of specificity.[6]

These characteristics of a personality frame stay fixed across periods of time under consideration in CDO_B (c). The qualification that x's personality frame "does not change between t_{n-1} and t_n" is needed because although we allow for changes of personality in a more distant past we need to exclude that the agent undergoes changes of her personality between times t_{n-1} and t_n. To find out if CDO_B is true about x, we only consider the possible worlds where x's personality frame is the same at t_{n-1} as x's personality frame in a at t_{n-1}. We don't want to say that "x could have done otherwise" is true about x if, in other possible worlds where x does otherwise, she has different skills or significantly different character, values, and beliefs, in particular, where x has different SODs. As we shall see, this removes some important problems.

Concerning CDO_B (d): Coherence of the FODs with the personality frame is the key step in the conditional analysis of CDO. We claim that x could have performed not-P out of her own free will if there is a possible world where x has different FODs that are coherent with her personality frame at t_{n-1} and x performs not-P at t_n, whereas all agent-external facts at t_{n-1} that were causally relevant to x's actual action at t_n are the same as in a. Coherence with one's personality frame (CPF) is characterized as follows:

> CPF. An action A of agent x is *coherent with x's personality frame F* iff performing the action A does not imply consequences that x can draw (using her instrumental reasoning abilities which are part of F) that contradict certain elements of F.

Condition CDO_B (d1) is a formal explication of what kind of changes internal to agent x are allowed in the possible worlds under consideration: namely, the

---

6 Since our account fixes all the abilities of a person in all ranges of specificity, it is consistent with both Whittle's (2010) and Kittle's (2015b) competing claims regarding what level of specificity of abilities is most relevant to free will.

changes in x's FODs that don't contradict CPF. Taken together with the fixity of F explicated in CDO_B (c), this restricts the counterfactual FODs to such FODs that are neither essential to F nor lead to actions that imply consequences that x can draw (using her instrumental reasoning abilities which are part of F) that contradict certain elements of F. Differences in world w that are mentioned in condition CDO_B (d1) reach all the way back to the Big Bang. These past differences are causally relevant to counterfactual FODs: x has a different FOD in w because a and w have different past histories that cause differences in the present states of these worlds, including the differences in x's FODs.

Condition CDO_B (d2) specifies that these differences must not affect those agent-external facts in the actual world at time t_{n-1} that were causally relevant to x's action at t_n. Otherwise, the counterfactual analysis would be trivialized and obviously unfree actions such as ones resulting from being forced by physical violence would come out as "free". However, these differences may affect those agent-external facts in the actual world at time t_{n-1} that were not causally relevant to x's action at t_n and may have further causal consequences in w at times later than t_n.

5 Discussion of the Backtracking Compatibilist Proposal

In this section, we will show how our account meets the intuitive desiderata by analyzing some pertinent examples. In the next section, we will discuss in detail the differences between our account and some other conditional accounts, and demonstrate that our account solves the problems that those accounts face.

First of all, we note that our account agrees with the classical conditional analysis on those examples that the latter gets right: those in which an external force prevents an agent from doing otherwise. For example, it follows from our account that if an agent is physically chained, he cannot move his arms even if he wanted to move them, because the causally relevant external conditions stay fixed in the counterfactual analysis of CDO_B.

Second, our account explains an important pre-theoretical intuition, according to which not everything that a person can physically do she can do in a free will sense of "can". Consider Jones: when a robber points a gun to his head and demands that Jones hand over his wallet, we do not want to say that Jones is free to do otherwise than obey the robber, even if there is a possible world where Jones refuses. If Jones, like most of us, values his life more than

his wallet, then his personality frame contains the essential FOD to preserve his life. According to our account, there is no possible world where Jones' personality frame F does not change and he refuses to hand over the wallet because we excluded all other possible differences, such as Jones' mishearing the threat, or having some form of hallucination, or being manipulated by neuroscientists, by the condition that everything except x's FODs that are not part of F in these possible worlds at t_{n-1} is the same as it is in a.

Third, our account answers an important objection to a standard conditional analysis of abilities raised by Inwagen (1983). Consider Smith, who is in a coma in a hospital. Van Inwagen observes that:

> The two propositions
>
> > Smith cannot get out of bed
> >
> > If Smith wanted to get out of bed, he would
>
> would seem both to be true, the former because he is in a coma, and the latter because, if he *did* want to get out of bed he wouldn't be in a coma. (Inwagen 1983, 119)

This objection is a problem for many other versions of classical compatibilism, but we think that there is a straightforward way to avoid this problem on our account, because being in a state of coma violates condition CDO_B (c), according to which x's personality frame is fixed in the possible worlds under consideration. But a person in a state of coma doesn't have a personality frame in the sense in which we understand this notion, because, while being comatose, the person is not moderately reasons-responsive, and at least temporarily lacks a capacity to form first- and second-order desires and volitions.

Fourth, our account solves the notorious red candy problem, dating back to Lehrer's (1968) example. The example is as follows:

> Suppose that I am offered a bowl of candy and in the bowl are small round red sugar balls. I do not choose to take one of the red sugar balls because I have a pathological aversion to such candy. [...] It is logically consistent to suppose that if I had chosen to take the red sugar ball, I would have taken one, but, not so choosing, I am utterly unable to touch one. (Lehrer 1968, 32)

The conclusion from Lehrer's thought experiment seems to be that a conditional analysis of abilities is bound to fail because in this case, it will give the implausible result that I can take the red candy if I decide/choose/want to, whereas intuitively I cannot choose a red candy because of my phobia. Our account, however, gives the conclusion that the person with a phobia cannot take the candy precisely because of his pathological aversion which is a part of his personality frame, so it has to be fixed in all of the possible worlds that we consider. It could be objected that we should also consider the possible worlds where something distracts the person with the phobia so that he forgets about his phobia at the critical moment. However, such a counterfactual distraction would be caused by a change in the agent-external facts at time t_{n-1} that were causally relevant to x's action at t_n, and that is excluded by condition (d2) of CDO_B. It could also be objected that some competing FOD can ultimately outweigh the aversion, so that a person takes the red candy after all, even if the phobia is included in F. However, the point of mentioning phobias in these kinds of counterexamples is precisely because they entail the inability of agents to form certain kinds of desires. Phobias distinguish what individuals cannot do from what they can do. Our account secures this intuition by including phobias and other kinds of irresistible psychological impulses into F. Consequently, any counterfactual FOD incompatible with the phobia is excluded by CPF. What is an irresistible psychological impulse and what is not (like weakness of will) is an empirical question, and the answer to this question determines whether some desires and fears should be fixed as elements of F.

Fifth, our account captures the pre-theoretical intuition about the cases in which agents freely perform some actions that are not very consequential for them. For example, we have a strong pre-theoretical intuition that we could have put on different clothes in the morning or ordered a different meal at the restaurant. The extreme case of choosing among inconsequential options is the so-called freedom of indifference: when an agent has to choose between two (or more) options and has no reason whatsoever to prefer one option over the other. This situation seemed problematic to those philosophers who thought that every choice must happen for a reason, i.e., be caused by a prior decision of the intellect (cf. Kenny 1973), but it is not problematic on our account, because any variation in the internal life of the agent, including the slightest unconscious biases or simply differences in neuronal activity will be enough for the agent to act otherwise.

Sixth, our account explains why we do not say that all animals have free will. According to a Humean style simple conditional analysis which says that *x* acted freely if he would have acted differently given different desires, all animals that have desires would also be free in their actions. It would follow, e.g., that a mosquito is freely stinging, because if it were not hungry but, rather, sleepy, it would do otherwise. However, a mosquito has no personality frame and therefore its actions are not free according to our analysis. For the same reason, primitive robots are not free according to our explication.

Finally, our account explains why human beings who have a rudimentary form of free will but have not yet developed a personality frame, such as young children, or who have a defective personality frame (for example, due to severe psychological disorders), do not qualify as free agents.

It may be objected that a satisfactory explication of freedom should also apply to situations in which a person changes her personality frame, but it is hard to say what freedom means in this case. Typically, an action that is involved in such a change violates some elements of the person's old frame but is in line with the person's new but not yet fully developed personality frame. So, what counts for the evaluation of an action as free or unfree in such a situation, the old (past) or the new (future) personality frame? If a person is manipulated by another person in a way that changes her personality, but after the change she considers herself free and her action is compatible with the new frame, then in which sense was this change unfree or free? We do not intend to develop a solution to this difficult but distinct problem in this paper; we postpone it to future work.

A final remark: Normally an agent's personality frame is not so strong as to determine her actions or first-order desires. In the exceptional case, however, in which someone does something as an immediate consequence of her personality frame, for example, regularly breathes, eats, and drinks (because the personality frame includes her desire not to die), then our present analysis implies that the person *is indeed not free* in regard to these actions. This sounds reasonable in the case of our example, but there are other cases where it does not seem so reasonable. Dennett (1984) draws our attention to cases where an agent's deeply held convictions make any alternative course of action inconceivable to the agent. According to Dennett, when Luther claimed "Here I stand; I cannot do otherwise" he might have been telling the truth, while still being free and responsible for the choice that he made. Regardless of whether this diagnosis really does apply to Luther on this occasion, it does seem plausible that sometimes there is only one way a person can act. There

are many deeply held convictions that make some courses of action inconceivable for certain agents. It might be this intuition of fixity of everything that is subjectively important for a person that brings some compatibilists to deny that having alternative possibilities is at all relevant to having free will. If one intends an analysis that does not make CDO a part of the explication in cases of free action to require a change of the personality frame, one has to change our defining condition by adding the following disjunct:

CDO_B*. "[...] or the action P follows already from the content of x's personality frame."

In this case, the modified definition of free action would be: x acted freely if either CDO_B or CDO_B* obtain. Whether CDO_B or CDO_B* (or something in between) is the better analysis of free action in a deterministic world is left here as an open question to be treated in future work.

6 The Advantages of the Backtracking Compatibilist Proposal

In the current free will debate one sometimes sees the contrast being drawn between the conditional analysis of abilities and the counterfactual possible world analysis, as if these two ways of analysis were mutually exclusive (Kittle 2015b, 101). We think that this understanding is mistaken since counterfactual possible world analysis is a way to provide a conditional analysis, as our paper illustrates. The contrast itself dates back to Lehrer (1976), who rejected conditional analysis and proposed his possible world analysis instead. However, it is important to note that what Lehrer rejected were the then available versions of a simple conditional analysis, which failed due to objections similar to those we considered in the previous section, including Lehrer's own red candy counterexample (1968), but not the very idea of finding a suitable conditional definition of free will. We think that our backtracking compatibilist proposal is a step towards such a definition, and we will now highlight how it differs from some influential versions of conditional analyses proposed by other authors.

Lehrer's (1976, 1990) possible worlds analysis states that a person is able to do otherwise if there is an accessible minimally different possible world where he does otherwise, and there is "no advantage" he has in that possible world as compared to the actual world. Lehrer's account of "advantage" was

persuasively criticized by Horgan (1977) and Kittle (2015b). These criticisms do not apply to our account, because CDO_B does not use the notion of advantage, but instead specifies in detail what is and what is not allowed to be different in the possible worlds under consideration. For this reason, our account is free from the difficulties that face the possible world analysis by John Campbell (1997), who develops Lehrer's notion of advantage.

Our account is also free from the problems that the new dispositionalist analyses of abilities face. Randolph Clarke (2009) has argued that the new dispositionalist accounts are vulnerable to objections similar to the red candy case, where an agent is unable to *A* because he is unable to try to *A*. He provides the following example:

> Suppose that on a certain occasion Bob formed an intention to wave to Cathy, but a momentary neural glitch made it impossible for Bob, on that occasion, to try to wave – he could not even begin to implement his intention – though he would have waved if he had managed to try. (Clarke 2009, 335–336)

Clarke argues that the new dispositionalism gives the implausible result that Bob was able to wave because he had a disposition to wave, which would have manifested itself if he had tried. We, however, answer that Bob was not able to wave, because by the condition below we fix all abilities of the agent at t_n:

> CDO_B (c). *x*'s personality frame agrees in *a* and *w* at all times until t_{n-1} and it does not change between t_{n-1} and t_n.

Due to the glitch, Bob lacks the ability to wave at t_n, because he temporarily lacks the proper causal basis *B* of this ability, namely, the normal functioning of his neural pathways. So he could not have waved at t_n.

Franklin (2011) argues that both Vihvelin (2004) and Fara (2008) succeed in providing dispositional accounts of narrow, or general abilities, but not of wide, specific abilities, or, as Franklin calls them, abilities with opportunities. This leads to an implausible claim that even externally constrained agents possess abilities to do otherwise:

> According to Vihvelin's analysis, free will is just a set of abilities, abilities are just (bundles of) dispositions, and dispositions are solely grounded in an agent's intrinsic properties. These claims prevent her from being able to appeal to the extrinsic features of

an agent (such as being tied to a chair) in order to explain why the agent is not free. (Franklin 2011, 97)

Our account does not have this problem, because the condition below excludes such changes in the agent's environment that prevent the agent from exercising her abilities:

CDO$_B$ (d2). w and a agree in all agent-external facts at t_{n-1} that were causally relevant to x's actual action at t_n.

Vihvelin (2013) proposes a modified account of narrow abilities that attempts to solve the problems raised by Clarke (2009) and Franklin (2011) by introducing a proportion of success cases:

(LCA-PROP-Ability) S has the narrow ability at time t to do R in response to the stimulus of S's trying to do R iff, for some intrinsic property B that S has at t, and for some time t' after t, if S were in a test-case at t and S tried to do R and S retained property B until time t', then in a suitable proportion of these cases, S's trying to do R and S's having of B would be an S-complete cause of S's doing R. (Vihvelin 2013, 187)

Kittle (2015a) argues that this modified account fails because it attributes to an agent abilities not relevant to free will. According to Kittle, Vihvelin's account has the following result: "When stood on the road miles from any water, I am such that *if I were in a test-case for my swimming abilities* and I tried to swim, then I would swim" (Kittle 2015a, 3031), but it would be wrong to conclude that I was free to swim there and then.

    Our account does not face this problem, because it specifies precisely which situations are the test-cases: those that fit the conditions of CDO$_B$.

7 Local Miracle Compatibilist Proposal

If the actual world a is deterministic, the local miracle strategy allows us to formulate the following conditions that have to obtain for CDO to be true about a person x in the actual world.

CDO$_M$. x could have done otherwise than P at t_n if x does not-P at t_n in some possible world w that satisfies the following conditions:

a. w is governed by deterministic laws that are identical to the laws of a except for the one local miracle mentioned in (b).

b. The pasts of w and a are identical until some past time t_m ($m < n$) at which a local miracle happens.

c. x's personality frame agrees in a and w at all times until t_{n-1} and it does not change between t_{n-1} and t_n.

d. (1) x's internal state at t_{n-1} in w differs from the corresponding internal state of x in a in regard to some FODs of x, in coherence with x's personality frame, where

    (2) w and a agree in all agent-external facts at t_{n-1} that were causally relevant to x's actual action at t_n.

Explications CDO_B and CDO_M differ in conditions (a) and (b), but are the same in (c) and (d).

Conditions CDO_M (c) and (d) provide forward-looking restrictions on what kind of miracles are allowed that are analogous to the backward-looking restrictions of the backtracking analysis and are needed for the same reasons. We have to exclude miracles that affect the agent-external facts in the actual world at time t_{n-1} (condition (d2)), since otherwise the counterfactual analysis would be trivialized. Moreover, also within the miracle account, we need to avoid an implausible conclusion that x could have done something that contradicts her personality frame. Indeed, imagine that Ann is sitting beside an open window in a high building and thinking about the fine day that awaits her. Can she freely jump out of the window for no particular reason? Of course, there is a possible world where she does precisely that due to a prior local miracle. But we would call such a situation a fluke, a random and unhappy incident, and not a free action of Ann's. The conclusion is that not only should the miracles leave x's actual personality frame intact, but also they should not bring about any consequences that are inconsistent with x's personality frame. This requirement is captured by condition CDO_M (d).

Condition CDO_M (d) also captures rationality requirements. Consider Jane: she is offered an apple and a pear and takes the apple. What has to be the case for the sentence "Jane could have taken the pear" to come out true? Presumably, a local-miracle compatibilist would not want to say that Jane could have taken the pear if there is a possible world where Jane decides to take the apple, but takes the pear instead, because of a prior miracle. This analysis would show that Jane could have done otherwise only if she had been irrational. It would not help much if a local-miracle compatibilist says

that Jane could have taken the pear if there is a possible world where Jane wants to take the apple, but decides to take the pear instead, because of a prior miracle. This would also be irrational. Condition CDO_M (d) provides us with the analysis that states that Jane could have taken the pear if there is a possible world where she forms a desire to take the pear, and takes it.

As the foregoing discussion shows, the conditions specifying which miracles are acceptable for CDO_M to be true about x are very similar to the conditions specifying which differences in the past states of the world are acceptable for CDO_B to be true about x.

Finally, our CDO_M analysis of abilities solves the problem for the local miracle compatibilism raised in Beebee (2003). Beebee argues that given the interpretation of abilities that can be reconstructed from Lewis (1981), there is no justification for the claim that the Weak Thesis is true, whereas the Strong Thesis is false because the possible world closest to the actual world where x does otherwise might be the possible world where x's act itself is a divergence miracle. Our local miracle compatibilist proposal does not have this problem, because condition CDO_M (b) specifies that divergence of w and a happens at some past time t_m earlier than t_n, so the local miracle cannot be an action of x at t_n, whereas conditions CDO_M (c) and CDO_M (d2) ensure that this divergence could not have been an action of x at some earlier time.

8 Indeterministic Compatibilist Proposal

In the previous sections we have shown how a philosopher can have a theory of free will compatible with physical determinism. Thereby we have defended free will against the objection based on the putative incompatibility asserted by the CA. However, there are strong (though not decisive) arguments, based on contemporary quantum physics, that physical determinism is probably false. Prima facie, indeterminism seems to be a much easier way to refute the CA and establish freedom in the sense of CDO, simply by denying premise 1. of the CA. However, free will sceptics argue that physical indeterminism poses other threats to free will, namely, the problem of irrationality and the problem of luck.

These problems arise for those libertarians who accept both the Fixity of the Past and the Fixity of the Laws, and deny Determinism. The problem of irrationality can be expressed as follows: in order to have free will at t_n, x must be able to do otherwise than P at t_n. x is able to do otherwise at t_n if there are worlds with the same laws of nature and the same past up to t_n where x does

otherwise at t_n. However, the same past up to t_n contains all of x's mental states and dispositions, including all of her first- and second-order desires, beliefs, deliberations and intentions (for short: deliberations) up to t_n. But if x is in fact justified in doing P at t_n as a result of her prior deliberations, then in other possible worlds she is acting irrationally when she performs not-P at t_n, as in our example with Jane who forms the desire to take the apple, decides to take the apple, but still takes the pear.

The standard libertarian response to this kind of worry consists in placing indeterminism not between the decision and the act, but between the desire and the decision, at some moment of deliberation when different motives and desires are being considered by the agent (Kane 1999, 2011a; Mele 2006; Ekstrom 2003).

This answer is good against the problem of irrationality. But it is not good enough to solve the problem of luck. Libertarians insist that a radically free act is never entirely determined by the past and the laws. No matter how firmly an agent decides to do something, how good her reasons are, and how strongly she wants it, she is free to do otherwise. This libertarian intuition has a troubling consequence when formulated in terms of possible worlds. Imagine that Mary, a libertarian agent, is considering whether she should cheat. She weighs pros and cons, thinks carefully, decides not to cheat, and acts in accordance with her decision. But she could have done otherwise, given precisely the same past up to t_n. Since everything about Mary is fixed right up to t_n after which either the situation w_1 where she doesn't cheat or w_2 where she cheats becomes actualized, it seems that if indeterminism obtains, then it is simply a matter of luck whether Mary cheats or not. And if some outcome is a matter of luck, it seems natural to say that the agent lacks control over this outcome, and therefore lacks free will in performing it. Some libertarian philosophers have devoted considerable efforts to address this problem (Kane 1999; Mele 2006). We claim that our indeterministic compatibilist account provides a solution to it, based on the conditional analysis of "can".[7]

If the world is indeterministic, the following conditions have to obtain for CDO to be true about an agent in this world:

---

7 We call this account "indeterministic compatibilist", and not libertarian, because, while it says that indeterminism is compatible with free will, it doesn't say that it is necessary for free will, whereas libertarian accounts do so. All three accounts we propose in this paper are versions of what Vihvelin calls "commonsense compatibilism", the position which maintains that "we actually have free will and that this is so regardless of the truth or falsity of determinism" (2013, 34).

CDO_I. x could have done otherwise than P at t_n if x does not-P at t_n in some possible world w that satisfies the following conditions:

a. w and a are governed by indeterministic laws that are identical.
b. The pasts of w and a are identical until some past time t_i during x's life span ($i < n$) at which x has spontaneously generated some counterfactual FOD in world w.
c. x's personality frame agrees in a and w at all times until t_{n-1} and it does not change between the time t_i and t_n.
d. 1. x's internal state at t_{n-1} in w differs from the corresponding internal state of x in a in regard to some FODs of x, in coherence with x's personality frame, where
    2. w and a agree in all agent-external facts at t_{n-1} that were causally relevant to x's actual action at t_n.

Explication CDO_I differs from CDO_B and CDO_M in conditions (a), (b), and (c), but is the same in (d). CDO_I (b) resembles CDO_M (b) because they both hold the past fixed until a divergence happens. There are two important differences between them: first, CDO_I (b) allows divergence of worlds' paths without miracles. Second, time t_i mentioned in CDO_I (b) is restricted to x's lifespan: x could not have generated a FOD before he came into existence, whereas time t_m mentioned in CDO_M (b) could be a time point before x is born. However, time t_i is not restricted to a short period between x forming a desire and x making a decision, as some libertarians argue in their solutions of the luck problem (Kane 1999, 2011a; Mele 2006; Ekstrom 2003). While CDO_I (b) does allow x's counterfactual FODs to be generated precisely in that time period (between x's desire and x's decision to act), it also allows for x's counterfactual FODs to be generated earlier. This provides a weak indeterministic position on an agent's free will, which does not require that in order for an agent to act freely an agent's choice must not be determined right up to the moment of the agent's making a decision. An agent will also be free even if he spontaneously generates a FOD sometime in the past, makes a plan in accordance with the FOD, and sticks to the plan. Thus, CDO_I seems to be a formal analysis capable of incorporating the intuition that sometimes we are really determined to do what we are doing because of the FODs we had some time ago, but we are nevertheless free because these FODs could have been different. However, what has to be required is that x's personality frame does not change in both worlds between the time t_i at which x spontaneously

formed the FOD causally relevant for his counterfactual FOD in world w at time t_{n-1} and time t_n. CDO_I provides a solution to the problem of luck similar to that of CDO_M without presupposing determinism. According to CDO_I, how x acts is not entirely determined by x's past and laws of nature. But it is not a matter of luck, because not every nomologically possible action could have happened with a corresponding probability, e.g., that the agent, instead of visiting his mother, could have killed his mother or ignored her for the next few months. Exactly that is afforded by our condition CDO_I (d) since it excludes all actions incoherent with x's personality frame. In other words, only actions coherent with x's personality frame are allowed. Therefore, it is no longer a matter of luck how x acts, although it is not determined either, because we are assuming indeterminism.

In conclusion, if physical indeterminism obtains, and spontaneous will-forming processes do indeed occur in our brains, then we claim that CDO_I is the correct analysis of alternative possibilities necessary for free will. On the other hand, if physical determinism obtains and spontaneous will-forming processes do not occur in our brains, then CDO_B or CDO_M can do the job. Either way, there is no reason to think that we need to know the truth about fundamental laws of physics before we can assert that some agents could have done otherwise.

9 Conclusion

We have provided a new account of free will, based on a conditional analysis of agents' abilities to do otherwise combined with sourcehood components. It allows alternative possibilities whether determinism or indeterminism obtains, and makes use of Frankfurt's psychological approach. Our proposal has three advantages:

1. It answers the objections against other versions of conditional analysis of "can" by demanding coherence of what one can freely do with one's personality frame, CPF. This allows us to analyze situations of coerced or irrational actions in an intuitively plausible way.
2. It is compatible with three metaphysical background assumptions:

   (i) determinism with backtracking
   (ii) determinism with local miracles and
   (iii) indeterminism.

3. It is immune to the consequence argument and also solves the luck problem.

Our account meets the intuitions behind the classical compatibilist approach, the sourcehood compatibilist approach, and the leeway libertarian approach. It is also not vulnerable to either the CA, which, according to a received opinion in the contemporary free will debate, is one of the most pressing worries for the compatibilists, or to the luck problem, which, according to another received opinion in the contemporary free will debate, is one of the most pressing worries for the libertarians. Therefore, it has the merits of both of these positions without having their drawbacks. Finally, our account of free will is naturalistic, because it is compatible with any answer that the fundamental physical theory can give to the question of determinism. Free will is real, and some agents have it, whether our world is fundamentally deterministic or not.*

Maria Sekatskaya
0000-0002-5381-2913
University of Düsseldorf
maria.sekatskaya@hhu.de

Gerhard Schurz
0000-0002-4107-9240
University of Düsseldorf
gerhard.schurz@phil.hhu.de

References

Austin, John Langshaw. 1961. *Philosophical Papers*. Oxford: Oxford University Press. Edited by J.O. Urmson and G.J. Warnock.

Ayer, Alfred Jules. 1954. *Philosophical Essays*. London: MacMillan Publishing Co.

Baker, Lynne Rudder. 2008. "The Irrelevance of the Consequence Argument." *Analysis* 68(1): 13–22, doi:10.1093/analys/68.1.13.

Beebee, Helen. 2003. "Local Miracle Compatibilism." *Noûs* 37(2): 258–277, doi:10.1111/1468-0068.00438.

BEEBEE, Helen, and Alfred R. MELE. 2002. "Humean Compatibilism." *Mind* 111(442): 201–224, doi:10.1093/mind/111.442.201.

BENNETT, Jonathan. 1984. "Counterfactuals and Temporal Direction." *The Philosophical Review* 93(1): 57–91, doi:10.2307/2184413.

BLUM, Alex. 2003. "The Core of the Consequence Argument." *Dialectica* 57(4): 423–430, doi:10.1111/j.1746-8361.2003.tb00281.x.

CAMPBELL, Joseph Keim. 1997. "A Compatibilist Theory of Alternative Possibilities." *Philosophical Studies* 88(3): 319–330, doi:10.1023/a:1004280421383.

CAPES, Justin A. 2019. "What the Consequence Argument Is an Argument for." *Thought* 8(1): 50–56, doi:10.1002/tht3.404.

CARLSON, Erik. 2000. "Incompatibilism and the Transfer of Power Necessity ." *Noûs* 34(2): 277–290, doi:10.1111/0029-4624.00211.

—. 2003. "Counterexamples to Principle Beta: A Response to (**?**)." *Philosophy and Phenomenological Research* 66(3): 730–737, doi:10.1111/j.1933-1592.2003.tb00287.x.

CLARKE, Randolph. 2009. "Dispositions, Abilities to Act, and Free Will: The New Dispositionalism." *Mind* 118(470): 323–351, doi:10.1093/mind/fzp034.

CRISP, Thomas M., and Ted A. WARFIELD. 2000. "The Irrelevance of Indeterministic Counterexamples to Principle Beta." *Philosophy and Phenomenological Research* 61(1): 173–184, doi:10.2307/2653408.

DENNETT, Daniel C. 1984. "I Could Not Have Done Otherwise – So What?" *The Journal of Philosophy* 81(10): 553–565, doi:10.5840/jphil1984811022.

—. 2003. *Freedom Evolves*. New York: Viking Press / Penguin Books.

EKSTROM, Laura Waddell. 2003. "Free Will, Chance, and Mystery." *Philosophical Studies* 113(2): 153–180, doi:10.1023/a:1023940209581.

FARA, Michael. 2008. "Masked Abilities and Compatibilism." *Mind* 117(468): 843–865, doi:10.1093/mind/fzn078.

FISCHER, John Martin. 1988. "Freedom and Miracles." *Noûs* 22(2): 235–252, doi:10.2307/2215861.

—. 1994. *The Metaphysics of Free Will*. Oxford: Basil Blackwell Publishers.

FISCHER, John Martin, and Mark RAVIZZA. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.

FRANKFURT, Harry G. 1969. "Alternate Possibilities and Moral Responsibility." *The Journal of Philosophy* 66(23): 828–839, doi:10.4324/9781315248660-2.

—. 1971. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68(1): 5–20, doi:10.2307/2024717.

FRANKLIN, Christopher Evan. 2011. "Masks, Abilities, and Opportunities: Why the New Dispositionalism Cannot Succeed." *The Modern Schoolman* 88(1-2): 89–103, doi:10.5840/schoolman2011881/26.

GINET, Carl. 1966. "Might We Have No Choice?" In *Freedom and Determinism*, edited by Keith LEHRER, pp. 87–106. New York: Random House.

—. 1980. "The Conditional Analysis of Freedom." In *Time and Cause: Essays Presented to Richard Taylor*, edited by Peter van Inwagen, pp. 171–186. Philosophical Studies Series n. 19. Dordrecht: D. Reidel Publishing Co., doi:10.1007/978-94-017-3528-5.

Gustafsson, Johan E. 2017. "A Strengthening of the Consequence Argument for Incompatibilism." *Analysis* 77(4): 705–715, doi:10.1093/analys/anx103.

Horgan, Terence E. 1977. "Lehrer on 'Could'-Statements." *Philosophical Studies* 32(4): 403–411, doi:10.1007/bf00368695.

Hume, David. 1748. *Philosophical Essays Concerning Human Understanding*. London: Andrew Millar of the Strand.

Inwagen, Peter van. 1975. "The incompatibility of Free Will and Determinism." *Philosophical Studies* 27(3): 185–199, doi:10.1007/bf01624156.

—. 1983. *An Essay on Free Will*. Oxford: Oxford University Press.

—. 2000. "Free Will Remains a Mystery." In *Philosophical Perspectives 14: Action and Freedom*, edited by James E. Tomberlin, pp. 1–19. Oxford: Basil Blackwell Publishers, doi:10.1111/0029-4624.34.s14.1. The Eight Philosophical Perspectives Lecture.

—. 2004. "Freedom to Break the Laws." In *Midwest Studies in Philosophy 28: The American Philosophers*, edited by Peter A. French and Howard K. Wettstein, pp. 334–350. Boston, Massachusetts: Basil Blackwell Publishers, doi:10.1111/j.1475-4975.2004.00099.x.

Kane, Robert H. 1999. "Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism." *The Journal of Philosophy* 96(5): 217–240, doi:10.2307/2564666.

—, ed. 2002. *The Oxford Handbook of Free Will*. Oxford Handbooks. New York: Oxford University Press, doi:10.1093/oxfordhb/9780195178548.001.0001. Second edition: Kane (2011b).

—. 2011a. "Rethinking Free Will: New Perspectives on an Ancient Problem." In *The Oxford Handbook of Free Will*, edited by Robert H. Kane, 2nd ed., pp. 381–405. Oxford Handbooks. New York: Oxford University Press, doi:10.1093/oxfordhb/9780195399691.001.0001. First edition: Kane (2002).

—, ed. 2011b. *The Oxford Handbook of Free Will*. 2nd ed. Oxford Handbooks. New York: Oxford University Press, doi:10.1093/oxfordhb/9780195399691.001.0001. First edition: Kane (2002).

Kenny, Anthony John Patrick. 1973. "Freedom, Spontaneity, and Indifference." In *Essays on Freedom of Action*, edited by Ted Honderich, pp. 89–104. London: Routledge & Kegan Paul.

Kittle, Simon. 2015a. "Abilities to Do Otherwise." *Philosophical Studies* 172(11): 3017–3035, doi:10.1007/s11098-015-0455-8.

—. 2015b. "Free Will and the Ability to do Otherwise." PhD dissertation, Sheffield: Department of Philosophy.

Lehrer, Keith. 1968. "Cans Without Ifs." *Analysis* 29(1): 29–32, doi:10.1093/analys/29.1.29.

—. 1976. " 'Can' in Theory and Practice: A Possible Worlds Analysis." In *Action Theory: Proceedings of the Winnipeg Conference On Human Action, held at Winnipeg, Manitoba, Canada, 9-11 May 1975*, edited by Myles BRAND and Douglas N. WALTON, pp. 241–270. Synthese Library n. 97. Dordrecht: D. Reidel Publishing Co.

—. 1990. *Theory of Knowledge*. 1st ed. Dimensions of Philosophy Series. Boulder, Colorado: Westview Press. Second edition: Lehrer (2000).

—. 2000. *Theory of Knowledge*. 2nd ed. Dimensions of Philosophy Series. Boulder, Colorado: Westview Press. First edition: Lehrer (1990).

LEWIS, David. 1979. "Counterfactual Dependence and Time's Arrow." *Noûs* 13(4): 455–476, doi:10.2307/2215339. Reprinted, with a postscript (Lewis 1986b), in Lewis (1986a, 32–51).

—. 1981. "Are We Free to Break the Laws?" *Theoria* 47(3): 113–121, doi:10.1111/j.1755-2567.1981.tb00473.x. Reprinted in Lewis (1986a, 291–298).

—. 1986a. *Philosophical Papers, Volume 2*. Oxford: Oxford University Press, doi:10.1093/0195036468.001.0001.

—. 1986b. "Postscript to Lewis (1979)." In *Philosophical Papers, Volume 2*, pp. 52–66. Oxford: Oxford University Press, doi:10.1093/0195036468.001.0001.

—. 1997. "Finkish Dispositions." *The Philosophical Quarterly* 47(187): 143–158, doi:10.1111/1467-9213.00052. Reprinted in Lewis (1999, 133–151).

—. 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press, doi:10.1017/CBO9780511625343.

MCKAY, Thomas J., and David JOHNSON. 1996. "A Reconsideration of an Argument against Compatibilism." *Philosophical Topics* 24(2): 113–122, doi:10.5840/philtopics199624219.

MELE, Alfred R. 2006. *Free Will and Luck*. Oxford: Oxford University Press, doi:10.1093/0195305043.001.0001.

MILLER, Dickinson S. 1934. "Free Will as Involving Determination and Inconceivable Without It." *Mind* 43(169): 1–27, doi:10.1093/mind/XLIII.169.1. Published under the name "R.E. Hobart" .

MOORE, George Edward. 1912. *Ethics*. Oxford: Oxford University Press.

SAUNDERS, John Turk. 1968. "The Temptations of 'Powerlessness' ." *American Philosophical Quarterly* 5(2): 100–108.

TAYLOR, Christopher, and Daniel C. DENNETT. 2002. "Who's Afraid of Determinism? Rethinking Causes and Possibilities." In *The Oxford Handbook of Free Will*, edited by Robert H. KANE, pp. 257–280. Oxford Handbooks. New York: Oxford University Press, doi:10.1093/oxfordhb/9780195178548.001.0001. Second edition: Kane (2011b).

TOGNAZZINI, Neal A. 2016. "Free Will and Miracles." *Thought* 5(4): 236–238, doi:10.1002/tht3.224.

VIHVELIN, Kadri. 1988. "The Modal Argument for Incompatibilism." *Philosophical Studies* 53(2): 227–244, doi:10.1007/bf00354642.

—. 2004. "Free Will Demystified: A Dispositional Account." *Philosophical Topics* 32(1–2): 427–450, doi:10.5840/philtopics2004321/211.

—. 2013. *Causes, Laws, and Free Will. Why Determinism Doesn't Matter*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199795185.001.0001.

Whittle, Ann. 2010. "Dispositional Abilities." *Philosophers' Imprint* 10(12).

Widerker, David. 1987. "On an Argument for Incompatibilism ." *Analysis* 47(1): 37–41, doi:10.1093/analys/47.1.37.