Two Intuitions about Free Will: Alternative Possibilities and Intentional Endorsement

Christian List and Wlodek Rabinowicz[1]

16 December 2011, this version 13 November 2014

forthcoming in *Philosophical Perspectives*

Free will is widely thought to require (i) the possibility of acting otherwise and (ii) the intentional endorsement of one's actions ("indeterministic picking is not enough"). According to (i), a necessary condition for free will is agential-level indeterminism: at some points in time, an agent's prior history admits more than one possible continuation. According to (ii), however, a free action must be intentionally endorsed, and indeterminism may threaten freedom: if several alternative actions could each have been actualized, then none of them is necessitated by the agent's prior history, and the actual action seems nothing more than the result of indeterministic picking. We argue that this tension is only apparent. We distinguish between actions an agent can possibly do and actions he or she can do with endorsement. One can consistently say that someone who makes a choice has several alternative possibilities, and yet that, far from merely indeterministically picking an action, the agent chooses one he or she endorses. An implication is that although free will can consistently require (i) and (ii), it cannot generally require the possibility of acting otherwise with endorsement.

1. The problem

Many of us have the following two intuitions about free will:[2]

> **Intuition 1: "Free will requires the ability to do otherwise."** An agent's action counts as free *only if* the agent could have acted otherwise.
>
> **Intuition 2: "Free will requires more than 'picking' some action in an un-determined way."** An agent's action counts as free *only if* the action stands in the right kind of relationship to the agent. It is not enough for the action to have been

[1] C. List, Departments of Government and Philosophy, London School of Economics; W. Rabinowicz, Departments of Philosophy, Lund University and London School of Economics. We gratefully acknowledge the hospitality of the Swedish Collegium for Advanced Study, where we wrote this paper. We are also grateful to Luc Bovens and John Broome for helpful discussion. Finally, we thank the Franco-Swedish Program in Economics and Philosophy and the Leverhulme Trust for supporting this work.

[2] For surveys of the debate on free will, see Kane (2002) and Fischer, Kane, Pereboom, and Vargas (2007).

> indeterministically picked from some set of alternative possibilities, for instance by some process of non-intentional randomization.[3]

For present purposes, we adopt a modal interpretation of the ability to act otherwise, according to which an agent can act otherwise if and only if it is *possible* for him or her to do so, in an appropriate sense of possibility.[4] According to Intuition 1, under this modal interpretation, a necessary condition for free will is indeterminism at the agential level.

> **Indeterminism at the agential level:** There are some points in time at which the agent's prior history has more than one possible future continuation, corresponding to different courses of action.[5]

Intuition 2 suggests a rather different picture of free will, according to which free will is tied not to indeterminism at the agential level, but to a special relationship in which the agent stands to his or her actions: a relationship of "agency", "authorship", or "endorsement"; we will make this more precise below. The juxtaposition of Intuition 1 and Intuition 2 leads to the following worry.

> **The worry:** Suppose we have indeterminism at the agential level, in accordance with Intuition 1. The agent's action at time *t* is one of several equally possible actions, given his or her prior history up to *t*. In particular, the modal relationship between the actual action and the prior history is no different from the modal relationship in which any of the other possible actions would have stood to that history. Each of these possible actions could have been actualized, and none of them – including the actual one – was specially supported, let alone necessitated, by the agent's prior history. It thus seems

[3] On the notion of "picking", see Stone (2014).

[4] A defence of the modal interpretation of the ability to do otherwise is beyond the scope of this paper; see List (2014). We further leave open the question of what precise sense of possibility is required. Our analysis is consistent both with (i) a metaphysical or physical notion of possibility, which free-will libertarians may adopt, and with (ii) an agential (more coarse-grained) notion of possibility, which some free-will compatibilists may favour, especially if they interpret free will as an emergent, higher-level phenomenon (e.g., Dennett 2003, List 2014). The idea of agential possibility has also received attention under the label of "agentive modalities". See, e.g., Maier (2011) for an analysis of "agentive modalities" on the basis of the concept of an agent's options. The main competitors of the modal interpretation of the ability to do otherwise are the conditional and dispositional interpretations: "the agent would do otherwise, if he or she chose (or tried) do so" and "the agent has a disposition to act otherwise in appropriate circumstances if he or she tries to do so", respectively. For the conditional interpretation, see Moore (1912, ch. 6) and Ayer (1954). The dispositional interpretation is a more recent proposal; see, e.g., Fara (2008).

[5] Like the possibility of doing otherwise, the notion of indeterminism at the agential level can be interpreted in different ways, depending on how we spell out the notions of "possibility" and an agent's "history". Although the more coarse-grained, agential interpretation of possibility (mentioned under (ii) in the previous footnote) is arguably most natural, what matters is that we use the same modal notions throughout the present paper.

> hard to attribute the actual action to the agent him- or herself; the action seems to have been merely indeterministically picked. According to Intuition 2, then, the agent's action cannot count as free.

In short, for an action to count as free, according to Intuition 1, it must not have been predetermined by the agent's prior history, whereas, according to Intuition 2, it must stand in some privileged relationship to that history: it must not have been merely indeterministically picked. Can there be any free actions under these constraints? The apparent tension between Intuitions 1 and 2 is one of the motivations behind the "hard incompatibilist" view, according to which free will in the conventional sense is compatible neither with determinism nor with indeterminism.[6]

The aim of this paper is to argue that there is no tension between Intuitions 1 and 2, and that there can be free actions in accordance with both intuitions. In order to reconcile the two intuitions, we must first explain what we mean by the "right kind of relationship" between an agent and his or her action, so that we can see whether such a relationship can coexist with the possibility of acting otherwise. Arguably, this relationship has both an intentional and a causal dimension: for an action to count as "more than just indeterministically picked", the agent must *intentionally endorse* that action, and he or she must have *causal control* over it. Both dimensions need some further spelling out. In this paper, we focus only on the first dimension – the intentional one – and hence speak of a relationship of *intentional endorsement*. Our central question, then, is whether an agent's intentional endorsement of his or her actions is consistent with the possibility of acting otherwise.

The second dimension – the causal one – raises a number of issues beyond the scope of this paper. Before setting it aside, however, we wish to note two points. First, even if an action is not predetermined, this does not imply that it is uncaused. There can be indeterministic causation, after all. Indeed, many processes in the world involve only probabilistic, as opposed to deterministic, causation; and so the possibility of acting otherwise need not conflict with causal control. Second, an agent's causal control over a particular action may depend on his or her intentional endorsement of that action: on a plausible account, an agent has causal control over an action *only if* the action is caused by the right kind of intentional state. So, even though we focus on intentional endorsement here, what we say is also relevant to the question of causal control.

---

[6] See, e.g., Pereboom (2001, 2005).

The key insight in support of our claim that the intentional endorsement of one's actions is consistent with the possibility of acting otherwise is that there is a distinction between (i) actions that an agent can *possibly* do, and (ii) actions that he or she can do *with endorsement*. This parallels a standard distinction in decision theory, namely between (i) actions that are possible (or "feasible") for an agent, and (ii) actions that are rational for him or her, so that the agent, if rational, would perform the latter (or some of them). With this distinction in place, one can consistently say that

- an agent who makes a particular choice has the ability to do otherwise, insofar as several actions are possible for him or her, none of which is predetermined, and yet that
- the agent intentionally endorses the chosen action and thereby stands in the required special relationship to it.

Of course, a full defence of free will would have to show, in addition, that the endorsement of that action plays the right causal role in the agent's choice, but, as noted, we set this issue aside here.[7]

Although Intuitions 1 and 2 can be reconciled in this manner, the reconciliation leads to an interesting and perhaps surprising qualification to our picture of free will: while we can consistently hold that free will requires the ability to act otherwise even if we accept Intuition 2, we cannot generally hold that it also requires the ability *freely* to act otherwise, where this is understood as the ability to act otherwise *with endorsement*. We develop this point in detail below.

2. A simple model

A formal model helps to present the architecture of our reconciliation of Intuitions 1 and 2.[8] We consider a world with a single agent and represent this world as a simple dynamic system. The system can be in a number of possible states, and its state evolves over time. We call the set of all possible states the *state space* and label it $\mathbb{S}$. It is best to interpret the elements of $\mathbb{S}$ as relatively coarse-grained, macroscopic states, say at the level of grain we find in psychological explanations,

[7] For a discussion, see List and Menzies (2009, forthcoming).

[8] Some of the present formalism – specifically the modelling of the ability to do otherwise in terms of branching in agential histories – draws on List (2014). Cf. Segerberg (1984), Belnap, Perloff and Xu (2001), and Belnap (2005).

rather than as very fine-grained, microscopic states, such as those we find in fundamental physics.[9] Further, let T denote the set of all points in time, where T is linearly ordered. For instance, T could be the set of all real numbers, the set of all natural numbers, or even just some finite set of natural numbers. No assumption about T apart from linearity is needed.

An *agential history* is a temporal path of the system through the state space, formally a function, $\mathbb{h}$, from T into $\mathbb{S}$, which assigns to each point in time, t in T, a corresponding state of the system, $\mathbb{h}(t)$ in $\mathbb{S}$. We write $\mathbb{\Omega}$ to denote the set of all agential histories that are possible relative to the laws governing the agential system, such as the laws of psychology. As is standard, subsets of $\mathbb{\Omega}$ can be interpreted as *propositions*, where a proposition is *true* in all histories contained in it and *false* in all other histories.

To express conditions such as determinism or indeterminism, we need to introduce the notion of *branching* in agential histories. It is helpful to use the notation $\mathbb{h}_t$ to denote the *truncated part* of history $\mathbb{h}$ up to time t, i.e., its initial segment up to that time. Formally, $\mathbb{h}_t$ is the restriction of the function $\mathbb{h}$ to the sub-domain of all points in time up to and including t. Note that $\mathbb{h}_t$ is itself a function, namely from the set of all points in time up to t into the set $\mathbb{S}$. Importantly, the truncated history $\mathbb{h}_t$ must not be confused with $\mathbb{h}(t)$, the agent's state at time t.

A *branching point* occurs in history $\mathbb{h}$ at time t if there exists another history $\mathbb{h}'$ which coincides with $\mathbb{h}$ up to time t and is distinct thereafter; formally, $\mathbb{h}_{t'} = \mathbb{h}'_{t'}$ whenever t' precedes t, and $\mathbb{h}_{t'} \neq \mathbb{h}'_{t'}$ whenever t' succeeds t.[10] A history is *deterministic* if it exhibits no branching point, and *indeterministic* otherwise. More broadly, our agential system is *deterministic* if all histories in $\mathbb{\Omega}$ are deterministic, and *indeterministic* otherwise.

We are now in a position to revisit our two intuitions about free will. Intuition 1 says that free choices require the ability to do otherwise. We interpret this as the claim that free choices can occur *at most* at branching points. Without branching points in agential histories, there cannot be

[9] Although our formal analysis could also be given for a more fine-grained state space, some interpretational issues become more cumbersome in that case.

[10] The fact that the *functions* $\mathbb{h}_{t'}$ and $\mathbb{h}'_{t'}$ differ whenever t' succeeds t does not rule out the possibility that the *states* $\mathbb{h}(t')$ and $\mathbb{h}'(t')$ might coincide for some such t'.

free will. But we can also see why this indeterministic requirement on free choices may seem to conflict with Intuition 2. Any branching point is characterized precisely by the fact that the same truncated history up to time *t* can have two or more distinct continuations. Nothing in the agent's truncated history privileges one such continuation over any other – at least not if we refer only to the elements of the model introduced so far. And so there is nothing that could make the actual continuation of the agent's history, as opposed to its counterfactual alternatives, count as standing in a special relationship to the agent.

Just as a light particle with the same past trajectory can indeterministically pass either through the left slit or through the right slit in the famous double-slit experiment, so an agent with the same past trajectory can indeterministically continue with one future trajectory or with another. The different continuations of the agent's history after time *t* stand in the same modal relationship – i.e., a relationship of possibility – to the agent's history up to *t*.[11]

To capture the idea that only some but not all possible continuations of the agent's history at a given time count as "intentionally endorsed" by the agent, an extension of the present model is needed. We need to introduce one further primitive notion: an *endorsement function*. Formally, this is a function

- whose domain is a set of truncated histories (typically, histories that are truncated at or before branching points, but histories truncated at other points need not be ruled out by definition),
- and whose co-domain is a set of sets of histories,

where

> any truncated history in the domain is mapped to the set of those continuations of that truncated history that are "intentionally endorsed" by the agent. Technically, we define a *continuation* of a truncated history $\mathbb{h}_t$ as a history $\mathbb{h}'$ such that $\mathbb{h}'_t = \mathbb{h}_t$.

[11] Note that we say all this against the background of indeterminism at the agential level. The modal notions used here refer to agential possibilities. As noted above, for the purposes of this paper, we do not take a stand on how the notion of agential possibility relates to any underlying notion of metaphysical or physical possibility.

Let us illustrate this notion. Suppose $\mathbb{h}_t$ is an agential history up to a particular point in time, such as a branching point at which the agent is faced with a decision. Provided that $\mathbb{h}_t$ is included in the domain of the endorsement function, $\mathbb{h}_t$ is then mapped by that function to the set of those continuations of $\mathbb{h}_t$ that the agent endorses. This will be a subset (not necessarily a proper one) of the set of all *possible* continuations of $\mathbb{h}_t$. We deliberately leave open the precise interpretation of "endorsement". This could be spelt out in a variety of ways, consistently with our formal framework, for instance by a substantive theory of intentional action. For a rational agent, in particular, "endorsement" can mean "rational support", so that the endorsed continuations of $\mathbb{h}_t$ are precisely those that are rationally supported in light of the agent's state in history $\mathbb{h}$ at time t. We say more about this later, when we discuss an illustrative decision-theoretic interpretation of endorsement.

Our definition of an endorsement function is consistent with a number of different *formal* properties of endorsement. For example, an endorsement function could map some truncated histories to the empty set. This allows for the possibility that some truncated histories have no endorsed continuation. Similarly, an endorsement function could map some truncated histories to a non-singleton set of continuations. This allows for the possibility that some truncated histories have more than one endorsed continuation. By contrast, whenever a truncated history is mapped to a singleton set of continuations, there is a unique endorsed continuation of that history.

It is further helpful to distinguish between two different cases of a non-singleton set of endorsed continuations of a given history. It may be that

(i) several possible continuations of $\mathbb{h}_t$ count as endorsed, but

(ii) they all coincide up to some *later* branching point t' and diverge only thereafter.

In such a case, we say that the set of endorsed continuations of $\mathbb{h}_t$ is *effectively singleton* (though not literally singleton), since there is only one endorsed *immediate future* after time t and up to time t'. By contrast, if (i) is true and (ii) is false (i.e., some histories in the set already diverge from

each other before a later branching point), we say that the set of endorsed continuations of h_t is *effectively non-singleton*.[12]

To ensure the generality of our framework, we do not impose any restrictions on the endorsement function. It should be clear that whenever the agent's truncated history at or before a particular branching point is mapped to a singleton or effectively singleton set of endorsed continuations, there is, in effect, only one immediate course of action that the agent can take *with endorsement*. The agent can still act otherwise at such a branching point, since there is more than one distinct future continuation of the agent's history (this follows from the definition of a branching point), but only one branch counts as endorsed. It follows that:

> **Observation:** Although at any branching point the agent can act otherwise, he or she may not generally be able to act otherwise *with endorsement*. Taking an endorsed action at a given branching point is accompanied by the ability to act otherwise with endorsement *if and only if* the set of endorsed continuations of the agent's history at that branching point is effectively non-singleton.

Crucially, the condition on the right-hand side of the last biconditional is not met in general.

3. The example of Martin Luther

An example that is frequently discussed in the literature on free will is that of Martin Luther.[13] When Luther was summoned to the Reichstag zu Worms in 1521 and was asked to renounce his critical views on the Roman Catholic Church, he reaffirmed his views, allegedly saying "here I stand; I can do no other". Does this mean Luther lacked free will on this matter? Most

---

[12] Our discussion of this complication is related to a more general problem: it is plausible that the alternatives between which an agent in a history h can choose at some time t are not fully specified histories that are continuations of h_t but mutually exclusive *sets* of such continuations. Depending on how fine-grained Ω is, the agent's powers of discrimination and capacities to control the future may be too limited to make the *fully specified* continuations available for choice. As discussed in more detail later, what is endorsed could therefore be a certain course of action, which corresponds to a set of continuations of h_t (all continuations compatible with taking that action). We count each of these continuations as endorsed when we apply the notion of endorsement to histories. The case in which the set of endorsed continuations is effectively but not literally singleton illustrates this: what the agent endorses here is a single course of action that fixes the future up to a certain point, but leaves several possibilities open thereafter. Sometimes endorsing a single action does not even manage to do this, e.g., if it can be performed in several different ways that are distinguishable in Ω.

[13] See, e.g., Dennett (1984) and Kane (2002, introduction).

commentators agree that, far from disavowing free will, Luther was actually taking responsibility for his actions, implying that these were a consequence of his character and commitments.

Should we interpret this example as showing that free will does not require the ability to act otherwise, thereby refuting Intuition 1? Daniel Dennett suggests this interpretation.[14] We think, however, that this would be too quick. A better interpretation of Luther's assertion would be that although it was not literally impossible for him to act otherwise, he could not have acted otherwise *without sacrificing his integrity*.[15]

Let us recast this example in the terminology of the previous section. When Luther arrived at the Reichstag zu Worms, he had in fact reached a branching point in his agential history. One continuation of this history would have involved a renouncement of his views; another – the one he actually pursued – a reaffirmation. Both continuations were literally possible, and what distinguished the second from the first was not its modal status, but the fact that Luther endorsed it, while he did not endorse the other continuation. Formally, Luther's endorsement function mapped his truncated history at the given time to a singleton (or effectively singleton) set: the set containing the history in which he reaffirmed his views. In short, Luther did have multiple possible futures at the time, but there was only one that he endorsed.[16]

These considerations suggest a distinction between two notions of a free action, which echoes a distinction familiar from related discussions, including Isaiah Berlin's discussion of freedom in a political, as opposed to metaphysical, sense (1958). According to the first, "negative" notion, an action counts as free if and only if the agent could act otherwise. This notion is sensitive only to Intuition 1. According to the second, "positive" notion, an action counts as free if and only if the agent endorses it. This notion is sensitive only to Intuition 2. Harry Frankfurt's famous argument that the ability to do otherwise is not necessary for moral responsibility may be described as relying on a positive notion of freedom.[17]

---

[14] See Dennett (1984).
[15] See also List (2014).
[16] For ease of exposition, we deliberately individuate histories in a very coarse-grained manner here, not differentiating between different possible versions of Luther's act of reaffirming his views (e.g., with respect to the precise choice of words, timing etc.). Luther himself invoked such a coarse-grained individuation of histories when he said "I can do no other". Formally, what matters is that Luther endorsed only those histories in which he reaffirmed his views, while there existed other possible histories that he did not endorse.
[17] See Frankfurt (1969).

Luther, we suggest, was free in both senses: he could have acted otherwise, and he did in fact endorse his action. But although he could have acted otherwise, the alternative was not one that he endorsed. If endorsement is necessary for freedom, then, although Luther could have acted otherwise, he could not have *freely* acted otherwise.

We do not think, however, that this last observation in any way challenges the claim that Luther acted out of his own free will in the given instance. While the exercise of free will may require both the ability to act otherwise and the endorsement of one's action (in accordance with Intuitions 1 and 2), it does not seem to require *the ability to act otherwise with endorsement*. The ability to act otherwise and the endorsement of one's action are arguably two separate dimensions of free will, which are not normally – and cannot generally be – entangled with one another.

4. An illustrative decision-theoretic interpretation of endorsement

We have been silent on how exactly to interpret the notions of agential possibility and intentional endorsement. Our aim has been to sketch a general architecture for thinking about the two intuitions about free will, without settling all the details. To show that it is possible to fill in the missing details, we now give an illustrative interpretation of our central concepts, drawing on standard decision theory.[18] This is not meant to be a definitive interpretation.[19] Our aim is merely to show that our framework is naturally compatible with established models of individual decision making.

In a standard model, an agent is faced with a choice between a number of different feasible *actions*. Each action can, in turn, have a number of different possible *outcomes*, where the actual outcome depends on which of a number of different possible *states of the world* obtains.[20] Now suppose the agent has a *utility function* over the set of possible outcomes, which assigns to each outcome the agent's utility under that outcome, and a *subjective probability function* over the set of possible states of the world, which assigns to each state its subjective probability from the agent's perspective. Then the agent can in principle calculate his or her expected utility for each of the feasible actions. If the agent is rational, he or she will perform an action that maximizes

[18] See, e.g., Savage (1954).
[19] Indeed, we think that standard decision theory is in need of refinement. See, e.g., Dietrich and List (2013).
[20] Formally, an action is a function from states of the world to outcomes, which maps each state to the outcome of taking the action in that state.

expected utility. More generally, an agent may be faced not only with a single choice, but with a sequence of choices, perhaps interspersed with choices by other agents or chance events. This may lead to an entire decision tree; we then speak of an "extensive-form" model.

A decision-theoretic model of this kind can be naturally embedded in the formal architecture we have suggested for thinking about free will. The agent's state at the time of a decision is given by the situation the agent is in at that time (which may include a previous sequence of decisions, previous chance events, and the agent's information set), together with the agent's utility and subjective probability functions. For simplicity, we assume that there is only a discrete number of time periods. Let $\mathbb{h}$ be the relevant history, and t the time of a decision. The occurrence of a "decision node" implies that there is a branching point in history $\mathbb{h}$ at time t. Different feasible actions correspond to different branches of history $\mathbb{h}$ at time t. Formally, a *branch* is an equivalence class of continuations of $\mathbb{h}_t$ that coincide initially and diverge at most after some later branching point. By corresponding to an available branch of history $\mathbb{h}$ at time t, each feasible action counts as *possible*. Yet, for a rational agent, only actions or action sequences that maximize the agent's expected utility count as endorsed. More precisely, the endorsement function maps the truncated history $\mathbb{h}_t$ to the set consisting of every continuation $\mathbb{h}'$ of $\mathbb{h}_t$ such that $\mathbb{h}'$ belongs to an expected-utility-maximizing branch at the present decision node and at all future decision nodes that occur along $\mathbb{h}'$. (At each decision node, expected utility is calculated according to the agent's utility-probability-function-pair, from his or her perspective at that decision node.) If we consider an entire decision tree, the set Ω of possible agential histories represents all *possible* paths through that decision tree, while the endorsed histories at any decision node represent the *rational* continuations of the decision path beyond that decision node.

In this way, decision theory accommodates both the possibility of acting otherwise and the existence of an endorsed action (or set of actions). While many actions may be feasible, only some count as rational (e.g., by maximizing expected utility). Summarizing this point:

> **Observation:** In a standard decision-theoretic model, multiple distinct actions are *feasible* for the agent and thus, in our terms, *agentially possible*, and yet only some (but typically not all) of the feasible actions are *rational* and thus, in our terms, *endorsed*.

In sum, an agent, according to standard decision theory, can have free will in both a negative and a positive sense: he or she (i) has the ability to act otherwise, provided there is more than one feasible action, and (ii) acts with endorsement if he or she is rational. The agent cannot generally *act otherwise with endorsement*, however, since the set of endorsed actions, unlike the set of feasible ones, need not contain more than one action. This completes the simple, illustrative interpretation of our framework.

5. Cases in which an agent's ability to act with endorsement is undermined

Our model shows that the ability to act otherwise is entirely compatible with the ability to act with endorsement, though not generally with the ability to act otherwise with endorsement. But the model can also be used to characterize cases in which, despite the presence of alternative possibilities, the agent's ability to act with endorsement is compromised. In cases of this kind, the agent is free in what we have described as a negative sense, but not in its complementary, positive sense. We now discuss four categories of such cases. All of them are theoretical possibilities, though depending on the precise interpretation of "intentional endorsement" some categories may remain empty.

5.1 No endorsed course of action at a particular branching point

In a well-behaved case, any truncated history $\mathbb{h}_t$ at a branching point will have at least one endorsed continuation. But this need not be so in general, unless our account of endorsement explicitly guarantees a non-empty set of endorsed continuations (as in the case of some decision-theoretic accounts that always identify at least one action as rational). In less well-behaved cases, the endorsement function might assign an *empty* set of endorsed continuations to $\mathbb{h}_t$. The agent would then endorse none of the continuations available at this branching point: all possible actions would be unacceptable from his or her point of view.

We can refer to such cases as *agential dilemmas*, in analogy to moral dilemmas, in which every available course of action is morally wrong. Sophie's choice from the well-known novel by William Styron (1979) may be an extreme example of an agential dilemma. In this example, Sophie is forced by an SS officer to choose which of her children should be saved and, by implication, which should be allowed to be killed. Clearly, she cannot – at least on a plausible

interpretation of her predicament – endorse any of the options, even though she must choose one in order to prevent the outcome that both of her children are put to death.

5.2 No fact of the matter as to which course of action is endorsed

A different kind of problem arises at branching points for which the endorsement function is *undefined* (another possibility that some accounts of endorsement may explicitly rule out). Formally, this happens if the relevant truncated history $\mathbb{h}_t$ is not included in the domain of the endorsement function. While in an agential dilemma it is true of every continuation that it is *not* endorsed, in the present case the question of whether a continuation is endorsed or not has no answer. There is no fact of the matter as to whether the agent endorses any given continuation of his or her history or not. Both the positive and the negative endorsement claims are indeterminate, given the state of the agent at time *t*. Choice situations outside the domain of what can be rationally adjudicated – if such situations exist – might give rise to such indeterminacies.[21]

5.3 Conflicts between different modes of endorsement

Yet another possibility is that an agent sometimes finds him- or herself torn between different reasons for or against the various actions, or between different ways of weighing them, without being able to arrive at a balanced view. Cases of incomparability or incommensurability, which are sometimes discussed in decision theory, fall into this category.[22] To represent cases of this kind, an extension of the present model may be required.

One modelling option would be to redefine an endorsement function as a mapping that assigns to each truncated history in its domain not a set of endorsed continuations but a *family* of such sets, with each set in the family being supported by one group of reasons, or by one way of weighing them. Now, for each such family, there are two possibilities: either the different sets in the family have some continuation(s) of the given truncated history in common, or they do not. In the first case, the agent is still able to make a "safe" choice, without having to resolve the underlying conflict of reasons: namely by opting for the continuation of his or her history that is supported by

[21] In a more complex model, we might allow an endorsement function to map each truncated history in its domain to a three-fold partition of the set of its possible continuations: one segment of the partition would contain those continuations that are definitely endorsed; another segment would contain those that are definitely not endorsed; and the third segment would contain those for which it is indeterminate whether they are endorsed or not.

[22] For further discussion of incomparability or incommensurability, see, e.g., Chang (2002) and Rabinowicz (2008).

each of the competing sets of reasons. In the second case, by contrast, the agent faces a problem similar, in some respects, to an agential dilemma: no continuation of the given history counts as unambiguously endorsed.

Another modelling option would be to keep the original definition of an endorsement function in place (i.e., to define it as a mapping from truncated histories to sets of endorsed continuations, rather than to families of such sets), but to allow that there might be more than one such function at work. The state of an agent who is torn between different reasons could then be described in terms of a "rivalry" between different endorsement functions – different criteria of endorsement – which pick out different and possibly disjoint sets of continuations of the agent's truncated history.

5.4 Intertemporal inconsistencies in endorsement

A well-behaved endorsement function may be expected to have the property of *intertemporal consistency*. Other things being equal, if a particular continuation of an agent's history is endorsed at an earlier branching point, it will continue to be endorsed at later branching points, so that there is no need for the agent to deviate from previously endorsed continuations of his or her history.

> **Intertemporal consistency:** For every truncated history $\mathbb{h}_t$ in the domain of the endorsement function, if $\mathbb{h}'$ is an endorsed continuation of $\mathbb{h}_t$, then, for every later point in time t' for which $\mathbb{h}'_{t'}$ is also in the domain of the endorsement function, $\mathbb{h}'$ remains an endorsed continuation of $\mathbb{h}'_{t'}$.[23]

For example, if the endorsement function is as in our illustrative decision-theoretic interpretation earlier on, then intertemporal consistency is satisfied. Recall that, in that example, the endorsement function maps any truncated history $\mathbb{h}_t$ at a decision node to the set consisting of every continuation $\mathbb{h}'$ of $\mathbb{h}_t$ such that $\mathbb{h}'$ belongs to an expected-utility-maximizing branch at the given decision node *and* at all future decision nodes that occur along $\mathbb{h}'$. The second conjunct ensures the satisfaction of intertemporal consistency.

[23] Formally, for any truncated history $\mathbb{h}_t$ in the domain of e, if $\mathbb{h}' \in e(\mathbb{h}_t)$ then $\mathbb{h}' \in e(\mathbb{h}'_{t'})$ for any time t' after t with $\mathbb{h}'_{t'}$ in the domain of e, where e is the endorsement function.

Although intertemporal consistency is an appealing property, there is no guarantee that an agent's endorsement function will always satisfy it. For this reason, there is room for a further way in which an agent's ability to act with endorsement may be undermined. Suppose that the agent's set of endorsed continuations of h_t is non-empty, but every continuation h' in it sooner or later ceases to be endorsed: at some later time t' at which a branching point occurs in h', h' will not be among the endorsed continuations of $h'_{t'}$. If the agent can already predict this at time t, he or she may well experience the choice situation as problematic: choosing an endorsed action now may preclude the choice of an endorsed action later.[24]

To summarize, we have identified four categories of cases in which the agent's ability to act with endorsement is undermined:

(i) agential dilemmas, in which the set of endorsed continuations of the agent's history at a particular branching point is empty;

(ii) indeterminate cases, in which the endorsement function is undefined at a given branching point, so that there is no fact of the matter as to which action the agent endorses;

(iii) unresolved conflicts between different reasons, where the agent is torn between different sets of endorsed continuations of his or her history or between different endorsement functions; and, finally,

(iv) intertemporal inconsistencies, where the agent can predict that none of the endorsed continuations of his or her history will continue to be endorsed in the future.

6. Endorsement as an intensional notion

So far, we have assumed that what is endorsed at any branching point in a given history is one or several *concrete histories*: (some of the) continuations of the truncated history at this point.

[24] Another issue should be mentioned here. Even when the endorsement function is intertemporally consistent, it is still important to define it also for branching points that lie "off any equilibrium path", so to speak, i.e., even for branching points that are reachable only if the agent at some earlier branching point chooses an action he or she does *not* endorse. Especially in game-theoretic contexts, such "off equilibrium" choices are important to consider, e.g., when one determines the backward-induction solution for an extensive-form game with perfect information by making predictions about what the players would choose at various future branching points, *if* they were to reach them.

However, this way of conceptualizing endorsement might be criticized for being too *extensional*. What we endorse, an objector might say, is not concrete histories but propositions that describe these histories and the actions performed in them in more or less detail. Such descriptions cannot be fully specific, given our limited abilities to grasp the future. From this *intensional* perspective, an endorsement function should map truncated histories not to sets of endorsed *histories*, but to endorsed *propositions*.

The first thing to note is that even our current extensional definition of an endorsement function admits a propositional interpretation. This is because truncated histories are mapped to *sets* of histories, which, in turn, can be interpreted as propositions. Recall that any subset of Ω, the underlying set of all possible agential histories, can be interpreted as a proposition, which is true in all histories contained in it, and false in all others. So, one might say, under the existing definition any truncated history is already mapped to an endorsed proposition.

Nonetheless, there are other, more general ways of constructing an *intensional endorsement function*. The general strategy is to define it as a function

- whose domain, as before, is a set of truncated histories,
- and whose co-domain is *either* (i) a set of propositions, *or* (ii) a set of sets of propositions,

where

> any truncated history in the relevant domain is mapped to (i) the proposition or (ii) the set of propositions that the agent "intentionally endorses".

Note that the two alternative ways of defining the output of the function – either as a single proposition or as a set of propositions – are formally distinct. Even if we reformulate the set of propositions in the second case as a single disjunctive or conjunctive proposition (which is possible if the set is finite), any such translation obviously involves some informational loss. So let us look at different ways in which we might specify an intensional endorsement function, and ask how these relate to extensional endorsement functions as defined earlier.

Consider first case (i), where the output of the intensional endorsement function for each truncated history is a single proposition. One might think that we have already covered this case. As noted,

an extensional endorsement function maps any truncated history to a set of endorsed histories, which in turn can be seen as a proposition: the proposition that is true in precisely those histories that are endorsed. So there appears to be a one-to-one correspondence between extensional and intensional endorsement functions, assuming the latter produce single propositions, not sets of propositions, as output.

The appearance of this one-to-one correspondence, however, is misleading. Since an extensional endorsement function maps any truncated history $\mathbb{h}_t$ to a set of continuations of $\mathbb{h}_t$, the proposition that it picks out is true only in continuations of $\mathbb{h}_t$, not in any other histories. This is a significant restriction. More plausibly, the agent's object of endorsement may be a proposition in general, which picks out *some* subset of $\mathbb{\Omega}$; this may intersect with the set of continuations of $\mathbb{h}_t$, but need not consist solely of such continuations.

We can then retrieve an extensional endorsement function from an intensional one as follows. For each truncated history $\mathbb{h}_t$ (the input of both functions), the set of endorsed histories (the output of the extensional endorsement function) is the intersection of the endorsed proposition (the output of the intensional endorsement function) and the set of possible continuations of $\mathbb{h}_t$. Since different propositions can have the same intersection with the set of possible continuations of $\mathbb{h}_t$, intensional endorsement functions stand in a many-to-one relation to extensional ones: different intensional endorsement functions can give rise to the same extensional endorsement function.

Note that this opens up two ways in which the agent's set of endorsed histories at a given time may be empty, so that an agential dilemma arises. It may be empty because the agent's endorsed proposition is an impossible one, represented by the empty set and thus false in all histories. Or it may be empty because the endorsed proposition, while represented by a non-empty subset of $\mathbb{\Omega}$, happens not to overlap with the set of continuations of the agent's truncated history. In this case, the endorsed proposition is not impossible by itself, but impossible in the given context.

Case (ii), where the output of the intensional endorsement function is a set of propositions, rather than a single proposition, has two sub-cases. They correspond to two ways in which we can interpret the set of endorsed propositions to which a given truncated history $\mathbb{h}_t$ is mapped. One possibility is to interpret this set as containing different equally endorsed action-propositions

among which the agent is indifferent; he or she seeks to enact one of them. Another possibility is to interpret the set of endorsed propositions as containing different goal-propositions that the agent seeks to make true simultaneously through his or her action; in this case, the agent seeks to act in such a way as to render all the endorsed propositions true.

Both sub-cases are generalizations of case (i). In both sub-cases, an intensional endorsement function still determines an extensional one, where the determination relation is many-to-one. In the first sub-case, the set of endorsed histories for any truncated history $\mathbb{h}_t$ is the set of those continuations of $\mathbb{h}_t$ in which at least one of the endorsed propositions is true. In the second sub-case, it is the set of those continuations of $\mathbb{h}_t$ in which all of the endorsed propositions are true.

Again, the set of endorsed histories may be empty. In the first sub-case, this happens *either* if no possible (i.e., non-empty) proposition is endorsed *or* if none of the endorsed propositions overlaps with the set of continuations of $\mathbb{h}_t$. In the second sub-case, there are four ways in which the set of endorsed histories may be empty. It may be empty because some proposition among the endorsed propositions for a truncated history $\mathbb{h}_t$ may be impossible (i.e., empty) or impossible in a given context (i.e., non-overlapping with the set of continuations of $\mathbb{h}_t$). Or it may be empty because the different endorsed propositions, while separately possible to realize, are mutually incompatible. Either they cannot be realized together at all, or they cannot be realized together in a given context. In the latter case, there are histories that instantiate all the endorsed propositions, but none of them is a continuation of $\mathbb{h}_t$.

6. Concluding remarks

We have developed a simple model to show that two familiar intuitions about free will can be formally reconciled: the intuition that free will requires the ability to do otherwise and the intuition that this ability alone is insufficient for free will, since free will requires more than "indeterministic picking". The central observation has been that there is an important distinction between *possible* and *endorsed* actions. Our model has offered conceptual resources for capturing both notions and for analysing their relationship to one another.

In line with other contributions to the literature, we have argued that there are at least two senses in which an action can be free: a "negative" sense, which ties freedom to the ability to do

otherwise, and a "positive" sense, which ties freedom to intentional endorsement. Our analysis shows that one can consistently hold that free will requires *both* the ability to act otherwise *and* the ability to act with endorsement, but that one cannot generally (but only in special cases) hold that free will implies the ability to act otherwise with endorsement. Recognizing this last point is crucial for resolving the apparent tension between the two rival intuitions we have started with.

While we have defended a formal model for thinking about free will, we have tried to remain as neutral as possible on the question of how to interpret the various elements of that model. In analysing an agent's ability to do otherwise, we have adopted a modal interpretation of this ability, but have not taken a stand on how exactly *agential possibility* is related to *physical* or *metaphysical possibility*. Both libertarians, who usually employ a physical or metaphysical notion of possibility, and those compatibilists who take agential possibility to be distinct from physical or metaphysical possibility should be able to use our model. And in defining an endorsement function, we have tried to stay neutral between different philosophical views on what endorsement consists in.

Our goal has been to discuss the relationship between two intuitions about free will from a structural perspective. This should allow our model to serve as a framework in which different substantive theories of freedom can be located and compared.

References

Ayer, A. J. (1954) "Freedom and Necessity", in *Philosophical Essays*, London (Macmillan), pp. 271-284

Belnap, N. D., M. Perloff, and M. Xu (2002) *Facing the Future: Agents and Choices in our Indeterminist World*, Oxford (Oxford University Press)

Belnap, N. D. (2005) "Branching Histories Approach to Indeterminism and Free Will", in Bryson Brown and Francois Lepage (eds.), *Truth and Probability: Essays in Honour of Hugues Leblanc*, London (College Publications), pp. 197-211

Berlin, I. (1958) *Two Concepts of Liberty*, inaugural lecture, Oxford (Clarendon Press)

Chang, R. (2002) "The possibility of parity", *Ethics* 112(4): 659-688

Dennett, D. (1984) *Elbow Room*, Cambridge (Cambridge University Press)

Dennett, D. (2003) *Freedom Evolves*, London (Penguin)

Dietrich, F., and C. List (2013) "A reason-based theory of rational choice", *Nous* 47(1): 104-134

Fara, M. (2008) "Masked Abilities and Compatibilism", *Mind* 117: 843-865

Fischer, J. M., R. Kane, D. Pereboom, and M. Vargas (2007) *Four Views on Free Will*, Oxford (Blackwell)

Frankfurt, H. (1960) "Alternate Possibilities and Moral Responsibility", *Journal of Philosophy* 66: 829-839

Kane, R. (ed. with introduction) (2002) *The Oxford Handbook of Free Will*, Oxford (Oxford University Press)

List, C. (2014) "Free will, determinism, and the possibility of doing otherwise", *Nous* 48(1): 156-178

List, C., and P. Menzies (2009) "Non-Reductive Physicalism and the Limits of the Exclusion Principle", *Journal of Philosophy* CVI(9): 475-502

List, C., and P. Menzies (forthcoming) "My brain made me do it: The exclusion argument against free will, and what's wrong with it", in H. Beebee, C. Hitchcock, and H. Price (eds.), *Making a Difference*, Oxford University Press

Maier, J. (2011) "The Agentive Modalities", working paper, Australian National University

Moore, G. E. (1912) *Ethics*, London (Williams and Norgate)

Pereboom, D. (2001) *Living without Free Will*, Cambridge (Cambridge University Press)

Pereboom, D. (2005) "Defending Hard Incompatibilism", *Midwest Studies in Philosophy* 29: 228-247

Rabinowicz, W. (2008) "Value relations", *Theoria* 74(1): 18-49

Savage, L. (1954) *The Foundations of Statistics*, New York (Wiley)

Segerberg, K. (1984) "Towards an Exact Philosophy of Action", *Topoi* 3: 75-83

Stone, P. (2014) "Non-reasoned decision making", *Economics and Philosophy* 30(2): 195-214