

Beyond the Luck Problem: Addressing Discrimination in Event-Causal Libertarianism

Maria Sekatskaya¹

¹ Heinrich-Heine Universität Düsseldorf

Funding: DFG, Research Group Inductive Metaphysics (FOR 2495), research grant SCHU 1566/11-2

Potential competing interests: No potential competing interests to declare.

Abstract

Event-causal libertarians argue that free will requires indeterministic neural processes. This paper introduces a new problem with this theory via thought experiments involving agents who satisfy Robert Kane's criterion of free will and moral responsibility to varying degrees. The paper argues that treating these agents differently based on the type of brain activity they exhibit constitutes unjust discrimination. This challenge is then extended to Mark Balaguer's account. The paper also explores the similarities and differences between the luck problem and the newly introduced discrimination problem. To solve the luck problem, one must demonstrate that neural indeterminism is compatible with moral responsibility. To address the discrimination problem, one must show that neural indeterminism is not just compatible with, but necessary for moral responsibility.

Keywords: discrimination problem; event-causal libertarianism; luck problem; moral responsibility; neural indeterminism.

1. Introduction

Event-causal libertarianism aims to reconcile free will with the scientific worldview and our intuitions about alternative possibilities. Event-causal libertarians share the following assumptions (Balaguer 2010; Ekstrom 2000; Franklin 2018; Kane 1996; 2011):

(1) *Indeterminism at the agential level* Agents can choose between alternative courses of action, where 'can' is understood as being incompatible with physical determinism. This assumption is often formulated as a requirement of alternative possibilities (AP) for the agents. Typically, incompatibilists argue that the sense of 'can' incompatible with physical determinism is the one that requires 'actualist' alternative possibilities, i.e., alternative possibilities given that the past and the laws of nature are held fixed right until the moment of choice¹.

(2) *Causal efficacy of the mental* Agents' intentions are among the direct causes of their actions.

(3) *Naturalism about the mental* All mental states and processes of an agent are identical to or realized by the physical

processes in the agent's brain.

In what follows, I will refer to these three theses as Event-Causal Libertarian Assumptions (ECLA). Robert Kane's account of free will (Kane 1996; 2007a; 2011; 2019) is one of the most influential versions of event-causal libertarianism, and it has had a major influence on subsequent ECLA theories. While these theories differ from Kane's original version, they share its most essential features.

The structure of this paper is as follows. This first section is introductory: it formulates the principles that all event-causal libertarians accept. In section two, I review some specific claims made by Kane's theory which are relevant to the objection I make to it later. In section three, I show that Kane is committed to the 'History Matters' (HM) principle: the historical criterion which asserts that in order to be morally responsible, an agent must have had enough self-forming actions in her past. I illustrate this principle with three thought experiments featuring agents that satisfy Kane's historical criterion to different degrees. In sections four and five, I argue that unless Kane provides an argument to the effect that the moral status of these agents is different, treating them differently will constitute unjust discrimination. In section six, I show how this challenge can be extended to Mark Balaguer's account. In section seven, I show the similarities and differences between the luck problem and the discrimination problem. I argue that even if we grant that event-causal libertarians can solve the luck problem, they still face the problem of discrimination. In section eight, I conclude the article by arguing that event-causal libertarians do not have an explanation why indeterministic neural processes are necessary for moral responsibility.

2. Ultimate Responsibility and Self-Forming Actions

Robert Kane explains free will without evoking an extravagant metaphysics of prime movers unmoved, noumenal selves, or agent-causes (Kane 1994; 1996). In this regard, Kane's theory is similar to many compatibilist accounts which provide different sets of naturalistic criteria of free will. These typically include certain psychological abilities together with the lack of external and internal constraints. According to Kane, compatibilist requirements are also important for libertarians (Kane 1996, 2011, 2019). However, they are not sufficient for free will because they do not allow either alternative possibilities (AP), understood in a traditional incompatibilist way, as specified in assumption (1) of ECLA, or ultimate responsibility (UR), defined as follows:

To be ultimately responsible for an action, an agent must be responsible for anything that is a sufficient cause or motive for the action's occurring.² (Kane 2007a, 14)

According to Kane, both AP and UR are necessary for free will. This shows that his concept of free will is inextricably linked with the concept of moral responsibility. For every action for which an agent is morally responsible, either she is ultimately responsible, or her responsibility is derived from the fact that this action is a deterministic consequence of a prior action for which she was ultimately responsible (Kane 1996, 39–40).

I would like to illustrate in some detail exactly how Kane's account fulfills ECLA:

(1) *Indeterminism at the agential level* is captured by the AP requirement. In this regard, Kane's account is incompatibilist. At least some of the agent's actions must be indeterministic: namely, 'self-forming actions', as will be explained presently.

(2) *Causal efficacy of the mental*. If an agent has formed an intention to do something and as long as she does not change her mind and no external factors intervene, this intention will cause the corresponding action. In this sense, the agent has control over her actions. Accordingly, in normal circumstances, the neural processes underlying intentions should not contain intrusions of indeterminism which would bring about unexpected and unwelcome results (Kane 1994, 43).

There is a tension between the 'no intrusions of indeterminism' and AP. AP presupposes that the agent can act one way or another given exactly the same past (i.e., the agent has 'actualist' alternative possibilities). But where exactly should the moment of divergence be without undermining the agent's control over her actions? If a mental state of the agent in exactly the same situation can cause different outcomes, then wouldn't one of these outcomes be merely an intrusion of indeterminism? Kane's solution to this problem is 'plural rationality', according to which the moment of indeterminacy, when different outcomes are possible given exactly the same past, is precisely the moment of intention-formation, and not any of the moments after the intention has been formed. An agent's action is plurally rational if it is caused by the agent's will which was "divided" between "incommensurable reasons sets", where "one powerful set of motives is pulling the agent in one direction, while another, different and 'incommensurable' (non-comparable) set of motives is pulling the agent in an opposing direction" (Kane 1994, 44). By endorsing one of these sets of reasons, an agent is performing a self-forming willing (SFW), i.e., a willing which settles what kind of person the agent becomes after the action (Kane 1996, 128). The outcome of an SFW is not determined, but it is not random either because whichever of these incompatible sets of reasons the agent endorses, they both were parts of the total psychological state of the agent prior to the moment of choice. Thus, they were both possible motives, and through an SFW one of them became the actual motive. Therefore, SFWs satisfy both UR and AP. The actions performed as a result of SFWs constitute 'self-forming actions' (SFAs) which form a subclass of the class of all free actions:

... indeterminism does not have to be involved in all acts done 'of our own free wills' Not all of them have to be such that we could have done otherwise in a manner that is undetermined, but only those choices or acts by which we make ourselves into the kinds of persons we are, with the wills we do have. I call such 'will-forming' choices or acts 'self-forming actions' or SFAs. (Kane 2019, 147)

(3) *Naturalism about the mental* Human decision-making is carried out by neural processes in human brains (Kane 1996, 28, 119–210, 131, 137–142).

In accordance with these conditions, Kane proposes the following empirically testable hypothesis: the indeterminate efforts of will which are involved in performing an SFW are realized by "complex chaotic processes in the brain, involving neural networks that are globally sensitive to quantum indeterminacies at the neuronal level" (1996, 130). Kane develops this suggestion in (1996, 130–142, 151; 2007a, 28–31, 33–38).

In the next three sections, I will explore the ethical consequences of this account by using thought experiments that feature specific cases of neural determinism and indeterminism.

3. *A Historical Principle of Moral Responsibility*

Let us now consider three thought experiments that will highlight some of the problems with Kane's account.

Carol: Imagine that a certain person, Carol, satisfying some set of compatibilist conditions (CC³), inhabits a physically indeterministic world. Carol has never performed an SFA, although she is able to perform SFAs. This ability should be understood as follows: the physical organization of Carol's brain is such that there can arise indeterministic neural processes required to perform an SFA, i.e., the indeterministic neural processes which are subjectively perceived as incommensurable sets of motives "pulling the agent in an opposing direction" (Kane 1994, 44). Once such a set of motives arise, Carol can perform an SFA by endorsing one of them, in the same sense of 'can' that Kane assumes. It just so happens that Carol has never exercised this ability. Carol has faced enough temptations in her life, but she has never resolved them through a truly indeterministic neural process, only through a deterministic one. The facts about her neural networks and her environment have always determined the outcome, and the quantum indeterminacies present in her world at the fundamental physical level have never influenced her decision-making at the neuronal level. Nevertheless, Carol's subjective experiences were phenomenologically indistinguishable from those of other people who performed SFAs in similar situations.⁴

Imagine that Carol is facing a choice between jumping into a river in order to save a drowning child or continuing to walk across a bridge. She hesitates a moment (It's so cold! The current is so strong!) but overcomes her fears and jumps. Her psychological processes in this particular situation are deterministic: at the moment she saw the child it was determined by her character and values that she would jump. According to Kane, Carol was not morally responsible for her action, because this action was neither an SFA nor followed from any prior SFA.

It is an important feature of Kane's theory that the lack of moral responsibility is not due to the absence of AP in this particular case but rather to the fact that there were no AP grounding SFAs in Carol's past, and consequently, no UR. If Carol had performed enough SFAs in the past, she would have been morally responsible for her action, even if her motives determined the outcome of her deliberation in this particular case. We can draw this conclusion from the discussion of the case of Luther. Daniel Dennett (1984, 133) argues that AP are not necessary for moral responsibility and supports this claim with the example of Luther, who might have been right when he claimed that he could do no other than break with the church in Rome, and still be responsible for his actions. Kane argues that this is not a reason to reject AP in principle but only in particular cases of responsibility derived from prior SFAs. Luther's responsibility without AP at the moment of action depended on the fact that he could have done otherwise in the past. This clarifies the connection between AP and UR in Kane's account: an agent can be morally responsible for an act even if this act is determined by the agent's character, but only if the agent had AP enabling his SFAs in the past (Kane 1996, 39–40)⁵.

Let's grant that Kane is right regarding Luther and the significance of SFAs in one's past. In this case, Carol, despite

satisfying CC, is not morally responsible because she fails to meet the criteria for UR.

Rachel: Imagine another person, Rachel, physically and psychologically indistinguishable from Carol except for one difference: when Carol and Rachel were three years old, they both faced a temptation to eat some candy they were not allowed to eat. They both overcame the temptation, but Carol overcame it via a deterministic psychological process, whereas Rachel overcame it via an indeterministic psychological process of an SFA. If Rachel is in the same situation as Carol, will she, unlike Carol, jump into the water freely? Will one SFA in her past make a difference?

In Kane's account, there must be some difference. The indeterminism in the past is relevant for the moral status of a person, starting from the earliest choices she makes in her childhood (Kane 1996, 181; 2007b, 174–75; 2011, 399). In providing an elaboration of Kane's version of event-causal libertarianism, John Lemos (2018) argues that although one act of indeterministic choice in early childhood does not confer on an agent the status of a truly free and responsible person, it does confer a status of a person free and responsible to a small degree. Lemos considers a scenario of a three or four year old child, Rosa, whose first SFA was to steal a cookie from a jar. Lemos argues that Rosa bears a minimal but real moral responsibility for this act and that this act becomes a foundation for her latter SFAs (Lemos 2018, 71).

Let us assume, for the sake of argument, that there is a minimal sufficient number N of such acts and that they all jointly influence the development of full moral responsibility of a grown-up free agent in a manner described in (Kane 1996, 181; Lemos 2018, 69–72). We can express this intuition in the 'History Matters' principle:

HM: A person is morally responsible for any action only if she has performed a sufficient number N of self-forming actions in the past.

Of course, it remains an open question whether the numerical value of N can be established independently of the context of each individual life history. Most probably, the value of N will vary depending on the significance of each particular SFA for the agent who performed it. A complete account of HM should cover different cases. It might be that Jim has performed four SFAs that were causally relevant to the formation of his character and that played an important causal role in many of his decisions, and Sue has performed forty SFAs but each contributed very little to the formation of her character and her future decisions.⁶ In such a case, Jim might have performed N SFAs and might have reached the status of a morally responsible agent, whereas Sue, while performing more SFAs than Jim, might still be far from reaching it. The HM principle, presupposed by Kane and Lemos, does not require that every morally responsible agent must have performed the same number of SFAs or that the same number of SFAs is necessary for being morally responsible for every action. Instead, it entails that unless an agent reaches a sufficient number N of SFAs, individually determined by her life circumstances, she does not satisfy UR.

After having introduced HM, I will introduce one more character who satisfies it, Linda. In the next section, I will demonstrate the problematic moral consequences of HM.

Linda: Imagine that another person, Linda, has performed N SFAs but is otherwise indistinguishable from Carol and Rachel. When Linda performs the same act in the same situation as Carol and Rachel, she is praiseworthy for what she

does because she satisfies both CC and HM.

4. *The Problem of Discrimination*

Assume that in a possible world w_1 there are many people like Carol (or C-people) and many people like Linda (or L-people). C-people are praised and blamed just as L-people are, but they deserve neither the praise nor the blame. According to HM, only L-people deserve praise and blame as well as just punishment⁷. But it is not known to anyone in w_1 who L-people and C-people are, and all of them are treated as L-people. If Kane is right, then our world might be identical to w_1 . Now imagine that a portable neural scanner (PNS) is invented in w_1 . Soon, PNSs are given to every newborn baby in order to record the baby's brain activity and report any malfunctioning to a doctor. These babies grow up to become a new generation of adults whose whole life of brain activities were recorded. It is now possible to discover who is an L-person and who is a C-person by analyzing the lifelong records of people's brain activities. A person is a C-person if all of her actions are retrospectively predictable with a high degree of probability from the recording of the neural activity prior to these actions. She is an L-person if some of her actions are not predictable in this way. An action is predictable if it can be predicted from the information about the neural activity NA_1 of a person who will perform this action. The neural activity NA_1 on which the prediction is based precedes by a sufficient amount of time the neural activity NA_2 underlying the intention to perform this particular action.⁸

One day, a man named Peter commits a crime. The evidence is conclusive, and Peter confesses that he has committed this crime. It seems that punishment is unavoidable. However, his lawyer, Paul, who studied w_1 versions of event-causal libertarianism at his law school, comes up with a brilliant idea: he will discover whether Peter's brain activity shows any evidence of the SFWs. He receives Peter's permission, requests the log of Peter's lifelong brain activity, has the best scientists analyze it, and receives the conclusion he hoped for; there is no evidence of any SFWs. It turns out that all of Peter's mental and physical acts can be retrospectively predicted from the recordings of the prior states of his brain. Paul argues in court that Peter did not commit the crime out of his own free will, and therefore does not deserve to be punished. The jury acquits Peter, persuaded by the philosophical arguments that Paul cites.

This case sets a precedent for w_1 . What should w_1 inhabitants do? There are two options:

4.1. *Unequal Treatment.* The inhabitants of w_1 can pass a law recognizing the different legal status of L-people and C-people. C-people will be absolved from legal responsibility whereas L-people will be held legally responsible for their actions. However, this seems morally problematic. Although not all cases of unequal treatment are necessarily bad, for example, most of us will agree that it is morally acceptable, and perhaps even obligatory, to treat minors and adults differently, that is, to treat people differently based on their age – a majority would not agree that it is morally acceptable to treat people differently based on their ethnicity, religion, or skin color. Would unequal treatment of L-classes and C-classes be the same as age-based unequal treatment or skin-color-based unequal treatment? If the HM principle is correct, then L-class and C-class based unequal treatment is like age-based unequal treatment: it reflects the objective truth about people's moral maturity. But many C-people would contend that this argument is the same as a similar

argument based on skin color. It is an objective fact that people have skins of different colors, but there is no reason to believe that some particular skin colors reflect moral superiority. Differences in ethnicity, social status, or religious convictions do not provide reasons for legal discrimination either. Similarly, although it is an objective fact that C-people and L-people differ with respect to the historical property of having had enough SFAs in the past, is there a reason to believe that this fact reflects the moral superiority of L-people over C-people, or that this fact provides a sufficient reason to grant C-people an indulgence for crimes and misdemeanors that L-people do not have? One can claim that this fact is just as arbitrary and morally irrelevant as many other facts that were at some point used to justify discrimination. The supporters of HM must show why their position is not a form of unfair neural-based discrimination.

One can object that Kane can respond that just as we consider it morally justifiable to discriminate based on age because we believe that people of a certain age are very likely to be full members of the moral community and thus morally responsible, it is also justifiable to discriminate based on whether individuals have had a sufficient number of SFAs, as SFAs are necessary for moral responsibility on Kane's view.

However, postulating moral significance of SFAs is part of Kane's theory and not something that everyone must accept independently of any proof. Kane explicitly assumes that SFAs are necessary for agents to satisfy UR, and according to his theory, having had enough SFAs (and satisfying HM) is necessary for any form of moral responsibility, whether ultimate or derived. Yet, there is a crucial disanalogy between SFAs and age. We can readily provide independent reasons why age matters for moral responsibility. It will be something typically included in compatibilist conditions (CC), such as rationality, reasons-responsiveness, understanding of good and bad, self-control, possession of second-order desires and volitions, and more. However, unless one already believes that having had N SFAs in one's past is morally significant, simply claiming it to be so because it follows from the UR principle assumed as an axiom in Kane's theory is not an independent reason to believe in it. The acceptance of UR itself is not argued for by Kane but assumed as intuitively necessary, and thus can be referred to as UR-intuition.⁹ My thought experiments challenge the HM principle, which is entailed by this intuition, by providing an intuition to the contrary: the intuition that discrimination against C-people would be unjust. In essence, my response to this objection relies on the "burden of proof" principle. I argue that the pre-theoretical intuition of C-people being unjustly discriminated against should be given at least some weight when compared to the equally pre-theoretical UR-intuition assumed by Kane's theory.¹⁰

My thought experiments do not aim to prove HM wrong but rather to demonstrate the need for a strong argument to support it. Perhaps HM is right, but since it leads to such drastic measures as unequal treatment of people who do not satisfy it while being no different in other respects from those people who do satisfy it, by the burden of proof principle I claim that HM itself needs a very persuasive defense. Such defense is lacking in the free will debate so far. What has been debated a lot is the relation between alternative possibilities (AP), free will, and moral responsibility.

Let us quickly review the AP, free will and moral responsibility interrelation in the debate, and see how HM gives this discussion a new twist:

1. AP are necessary for free will (accepted by libertarians and classical compatibilists).
2. Free will is necessary for moral responsibility (accepted by libertarians and classical compatibilists).

3. x did not have AP.

Therefore,

4. x is not morally responsible.

Dennett's Luther example aims to show that intuitive ascriptions of moral responsibility can sometimes be made even in cases where an agent lacks AP. In other words, that 3. can be true while 4. is false. If AP are necessary for free will, and free will is necessary for moral responsibility, the lack of AP entails the lack of moral responsibility. Therefore, any example where an agent lacks AP but is intuitively morally responsible constitutes a counterexample to the short argument above, and either 1. or 2. must be rejected.

Kane's answer to this counterexample is to shift the acts of free will to the past, thereby implying HM. Consequently, Kane still thinks that 1. is true in general, as he says, for example, in (1996, ch. 6), but it is not true for each particular act. In other words, AP are not a necessary condition for each act of free will. What is necessary, according to Kane, is UR, as was explained above.

However, my thought experiments show that while 1. might seem intuitive enough to accept it without argument, HM is not equally intuitive. When a person is confronted with HM, she might feel that what matters for moral responsibility are psychological and other factors included in CC, the general libertarian understanding of AP (that the agent is able to act otherwise given the same past by forming an SFA), but not historical facts about this agent's neural processes in the past. In other words, since HM leads to such a drastic and morally loaded policy as Unequal Treatment, it must be very persuasively argued for by its proponents. Unless such an argument is given, Unequal Treatment is not justified and should not be accepted¹¹.

4.2. Concealment. If HM is true, the moral difference between L-people and C-people is real. However, if Unequal Treatment cannot be decisively justified, another possible solution would be to conceal the truth: to ban the usage of PNSs or at least to ban the storage of the lifelong recordings of PNSs by implementing a law that demands that the recordings should be only kept for a limited amount of time. However, this policy would be hard to defend. How can the ban of the usage of PNSs be justified if the only reason to implement it is not some deficiency or danger of the technology itself but the need to conceal the truth? Moreover, any possible justification would face ethical problems. The rationale behind Concealment is that the HM principle is true: there is indeed a moral difference between L-people and C-people. But if this difference is real, it would be ethically wrong to conceal it because L-people will never get the praise and blame they deserve (there will always be a doubt that they might be C-people, after all!) and C-people will never be justly blamed, because whenever they are blamed, it will be ethically wrong to blame them, since they did not perform the blameworthy act freely. Moreover, if free will is necessary for legal responsibility, punishing C-people would be unjust since most libertarians, including Kane, agree that just punishment presupposes free will (Kane 1996, 83). I conclude that concealment is not an ethically acceptable solution.

5. *C-People's Discrimination is a Problem for Event-Causal Libertarians*

5.1. A Problem for Robert Kane. These thought experiments are problematic for Kane's event-causal libertarianism for at least two reasons. First, an empirical worry: world w_1 could, for all we know, be the world we will live in if PNSs are invented in the future, and there seems to be nothing contradictory in the possibility of PNS. Arguably, in our world, some brain scanning technologies are already coming close to it. Therefore, if HM is a problem for w_1 , it could become our problem, too.

In response to this, one might argue that in our world, as a matter of nomological necessity (i.e., as a consequence of natural law) all people perform N SFAs by the age they reach legal adulthood. However, this natural law, apart from seeming to be a rather strange coincidence with our legal practices, will not solve the discrimination problem. Although this law will guarantee that in our world there are no C-people to be discriminated against, the existence of this law will fall short of explaining whether such discrimination would be morally right or wrong. However, a theory of free will must provide an answer to this question.

And this is the second problem for Kane's account. As I have shown, this account entails HM, and HM entails a possibility that C-people and L-people exist. Consequently, Kane must either give arguments in favor of unequal treatment or agree that there are no morally relevant differences between C-people and L-people. The latter, in turn, will mean that HM is false and the historical facts about whether any person has performed enough SFAs are morally irrelevant.

It can be objected that some particular decisions in a court of law, considered in section four, are not a proper counterexample to a philosophical account of free will. Indeed, Kane makes clear that his account is metaphysical, and not based simply on some sort of legal practice¹²:

The free will issue is metaphysical [...] We should therefore expect debates about the significance of free will to delve rather deeply into the metaphysical depths—and that is what they must do. Such debates are not simply about practical questions that arise in everyday life or law courts concerning responsibility, blameworthiness, and desert. Compatibilist accounts of free and responsible agency are often adequate to such questions. But debates about free will go beyond the practical concerns of the law courts and everyday conflicts. (Kane 1996, 91–92)

To this I reply that I am not arguing about what a particular judge or juror should decide in some particular case, but about what position would be philosophically motivated in the light of the empirical data which is supposed to obtain in my thought experiment. Kane himself admits that empirical data are relevant to our philosophical account of free will. So, if we think that free will is relevant to moral responsibility and if moral responsibility is relevant to justice then my thought experiments should be addressed by those ECLA theorists who want their theories to track the same kind of moral responsibility that should be presupposed, even if not philosophically analyzed, by a qualified judge.

5.2. A Problem for ECLA. Can any general conclusion be drawn for the whole approach of event-causal libertarianism? At first glance, it might seem that the challenge I have brought up is only a challenge for Kane's particular version of event-causal libertarianism, as well as for Lemos' elaboration of this account. Indeed, my criticism is based on the moral significance of SFAs and their historical nature, as well as on Kane's claim that sometimes an agent can be morally

responsible for a determined act if this act was a consequence of this agent's prior SFAs. This is already an interesting result, because Robert Kane's version of event-causal libertarianism is one of the most influential and widely discussed naturalistic libertarian theories (cf.: Palmer 2014). But the underlying problem which gives rise to this challenge is more general than that. The problem is that it is unclear why the fact that some neural processes of an agent are indeterministic is relevant for a seemingly unrelated issue of whether this agent is morally responsible. For many years, compatibilists have presented accounts aiming to clarify the relationship between certain psychological features of an agent, certain characteristics of her action, and the agent's moral responsibility. Libertarians have often argued that these accounts, although they manage to capture some necessary conditions for moral responsibility, are not the whole story, because they fail to capture a necessary metaphysical condition of indeterminism. Naturalistically minded libertarians have to locate indeterminism somewhere in the agent's neural processes. The challenge that my thought experiments pose is the challenge to justify the moral significance of neural indeterminism, which, according to the event-causal libertarians, allows the physical realization of the metaphysical condition of indeterminism.

There are several influential versions of event-causal libertarianism which differ from each other and from Robert Kane's version while subscribing to ECLA (Balaguer 2010; Ekstrom 2000; 2011; Franklin 2018). It is impossible to do justice to all of them in one paper and to show in detail how my thought experiments create a challenge to every particular version. However, in the next section, I will show how this challenge can be extended to Mark Balaguer's account. I choose Balaguer because his theory is partly similar to Kane's, but also different from it in important respects. I aim to demonstrate that event-causal libertarians need to justify the moral significance of neural indeterminism even if they do not openly postulate such historical conditions as UR.

6. Is L-Freedom Necessary for Moral Responsibility?

Mark Balaguer accepts ECLA and claims that indeterminism is necessary for L-freedom (this term is introduced by Balaguer as, approximately, a replacement for 'free will according to libertarians'¹³) and is realized in the neural processes of agents. Balaguer formulates precise criteria about the way that indeterminism must operate in order to enhance free will. It must be present in situations of 'torn decisions,' which are characterized by their phenomenology:

A torn decision is a decision in which the person in question (a) has reasons for two or more options and feels torn as to which set of reasons is strongest, that is, has no conscious belief as to which option is best, given her reasons; and (b) decides without resolving this conflict—that is, the person has the experience of “just choosing”. (Balaguer 2010, 71);

However, not all torn decisions but only wholly undetermined ones are necessary for L-freedom:

... to say that a torn decision is wholly undetermined at the moment of choice is to say that the moment-of-choice probabilities of the various reasons-based tied-for-best options being chosen match the reasons-based

probabilities, so that these moment-of choice probabilities are all roughly even, given the complete state of the world and all the laws of nature, and the choice occurs without any further causal input, that is, without anything else being significantly causally relevant to which option is chosen. (Balaguer 2010, 78).

Whether any particular torn decision is wholly undetermined is an empirical question: it can have the “torn” phenomenology without the proper physics, namely, if neural processes which realize this decision are not truly indeterministic (Balaguer 2010, 69). Balaguer agrees with Kane that an agent has plural authorship and control over her actions when she exercises free will in cases of torn decisions (Balaguer 2010, 83). However, he disagrees that a conflict of incommensurable motives and efforts of will are necessary for a freely willed act (Balaguer 2010, 74). In this respect, torn decisions are different from SFAs, and they are also much more numerous. They actually happen all the time; not only when we choose between different job offers, vacation destinations, or between a selfish and an altruistic act, but also when we choose between a vanilla and chocolate ice-cream (Balaguer 2010, 72, 74, 105; 2019). A torn decision is characterized neither by its moral significance nor by the formative influence it has on the agent’s life, but by its phenomenology and physics. If a decision is experienced as torn and is wholly undetermined, then this decision is L-free.

I will use the thought experiments introduced in previous sections to show that the connection of L-freedom and moral responsibility is problematic. Since it is an empirical question whether any of the torn decisions of an agent are causally undetermined in the appropriate way, and given that even in the life of one particular agent not all of the torn decisions are undetermined (Balaguer 2010, 77, 103, 105, 174–175; 2019), it is possible that although an agent is capable of truly indeterministic torn decisions, that agent has never made such decisions. Here we can use the same character, with slight modifications, in order to fit Balaguer’s theory. Imagine that Carol* is psychologically and physiologically indistinguishable from agents with L-freedom, with the following exception: all of her decisions were realized by deterministic neural processes. Sometimes her decisions were made with certainty because they were obviously the best choice in those situations, i.e., they were ‘strongly rational’ in Balaguer’s terminology (Balaguer 2010, 114). Sometimes her decisions were phenomenologically experienced as torn, but they were physically determined: some kind of deterministic neural process settled which of the tied options was chosen. Consequently, Carol* is not L-free, because

*a **person** is L-free if and only if she makes at least some decisions that are such that (a) they are both undetermined and appropriately nonrandom [...], and (b) the indeterminacy is relevant to the appropriate nonrandomness in the sense that it increases or procures the nonrandomness. (Balaguer 2010, 121–122)*

However, she is capable of wholly undetermined torn decisions in the sense that her world is fundamentally indeterministic and her decision-making can proceed via an indeterministic neural process. Now, a familiar story:

Carol* is walking over the bridge, sees a drowning child, shortly deliberates, jumps into the water, and saves the child. Before jumping, she has made a torn decision to jump that was not wholly undetermined. Therefore, this decision and the subsequent act were not L-free, and none of her other acts and decisions were L-free.

Is Carol* morally responsible for her action? Although Balaguer does not claim that L-freedom is necessary for moral

responsibility (Balaguer 2010, 55–63), the conclusion that Carol* is morally responsible while not being L-free will be a problem for his account because it will put serious pressure on his claim that L-freedom is valuable and worth wanting (Balaguer 2010, 98, 107, 112–113). If L-freedom is not necessary for moral responsibility, then this is not the sort of freedom that most philosophers of free will are looking for. Moreover, Balaguer argues that there are at least *prima facie* reasons to think that our pre-theoretical concept of moral responsibility presupposes that agents have L-freedom (Balaguer 2010, 62). Therefore, from the best interpretation of Balaguer's version of event-causal libertarianism, it follows that Carol* is not morally responsible.

Enter Rachel*. She is identical to Carol*, except that when she sees a drowning child and makes a torn decision to jump into the water, the neural processes underlying this decision are wholly undetermined. This is the first L-free decision and L-free action in her life.

Is Rachel* morally responsible for this action? In the case of Rachel, the answer followed from Kane's and Lemos' explanations of the development of moral responsibility in children. Since Balaguer does not give such an explanation, we can try to see what the most plausible answer to this question would be given Balaguer's account.

6.1. No Responsibility: Rachel* is not at all morally responsible for this action. This seems to be wrong. If Rachel* does not bear even the most rudimentary moral responsibility for her first L-free act, the connection between moral responsibility and L-freedom is lost. If the first L-free act does not bring even minimal moral responsibility, why should the second (or the third, or, in general, N_{th}) L-free act bring any? If moral responsibility is connected to L-freedom, then either it is fully present immediately after the first L-free decision, or it is gradually achieved via a number of L-free decisions, starting with the first such decision. The claim that moral responsibility is absent until the $N-1_{th}$ decision and present after the N_{th} decision is not credible.

6.2. Partial Responsibility: Rachel* is morally responsible for this action to a certain degree. It is easy to see that this answer immediately leads to a formulation of the HM* principle:

HM*: A person is fully morally responsible for any action only if she has performed a sufficient number N of L-free acts in the past.

If Balaguer opts for the solution 6.2., this will show that his account, as well as Kane's and Lemos' accounts, presupposes the moral relevance of indeterminism in the past history of the agent. HM* principle is very similar to the HM principle, it varies only due to the difference between the SFAs and L-free acts. Since both SFAs and L-free acts are empirically grounded in an agent's indeterministic neural processes, a similar story about a possible world with PNSs can be constructed for the HM* principle. The conclusion is the same: a proponent of the HM* principle must explain why the historical fact that enough of a person's torn decisions in the past were wholly undetermined is relevant to the moral status of this person in the present. If such explanation is not given, the challenge of discrimination will arise for Balaguer's as much as for Kane's and Lemos' accounts.

6.3. Full Responsibility: Rachel* is fully morally responsible for this action. The justification for this answer can be the

following. Rachel* satisfies CC and she has performed an L-free act. Hence, she also satisfies Balaguer's conditions of free will and, presumably, moral responsibility. What reason is there to claim that she is not morally responsible?

The reason is that it would be unfair to make a significant moral distinction between Carol* and Rachel* *Ex hypothesi*, these two women are phenomenologically indistinguishable and they make the same choice for the same moral reasons. If one of them deserves moral credit (praise for her act), and the other does not deserve this moral credit, then the difference between them must be morally significant. Moreover, the fact that this difference exists must somehow depend on them; there must be something that one of them has done and the other has not done which explains why only one of them deserves moral credit¹⁴. In particular, if Carol* finds out that she does not deserve praise for her altruistic act, whereas Rachel* does, Carol* can protest: "But it is unfair! What could I have done differently in order to deserve the praise? I have done all I could!" Of course, given Balaguer's theory, we can answer: "There is a significant difference between you and Rachel*. When Rachel* was making her torn decision, her neural processes were wholly undetermined, so it was an L-Free decision, but when you were making your torn decision, your neural processes were not wholly undetermined. It was deterministic. But it could have been indeterministic: If M neurons rather than N neurons fired in a particular brain region, then your decisions would have been truly indeterministic, but N neurons in a particular brain region did fire, so your decision was deterministic." This answer, of course, will not satisfy Carol*: "That's all very well, but what could I, Carol*, have done differently in order to make this difference? What have I done wrong that Rachel* has done right?" The answer to this question, I submit, is "Nothing." An agent cannot control her neural processes on the level of individual neurons. There was nothing that Carol* could have done differently in order to be more like Rachel*. She did the same thing as Rachel*: she deliberated whether to jump or not and decided to jump. It just so happened that her deliberation process was deterministic and not indeterministic, and it would be unfair to treat her differently for a difference that was not her fault. Microlevel differences in how one's neural processes are realized is not something that the agent can control. Therefore, objection against treating agents differently based on how their neural processes are realized (uncontrollably by those agents) can be seen as a version of the discrimination objection. This difference in treatment must be persuasively argued for, and not just postulated as following from some theory.

7. The Challenge for Event-Causal Libertarianism

The thought experiments discussed above show that event-causal libertarians who accept ECLA face the challenge of explaining why adding indeterminacy to the neural processes which are causally relevant to an act changes the moral status of a person who performs this act. The challenge is to demonstrate how indeterminism in neural processes alone, without any further metaphysical additions, can help to bring about moral responsibility.

It can be objected that the problem of discrimination is just a version of the luck problem that has been widely discussed in relation to event-causal libertarianism. However, while being somewhat similar, these two problems are different in important respects. The luck problem can be formulated as follows: if an agent performing a free act has alternative possibilities until some moment t (the moment of choice or some other moment preceding the moment of choice¹⁵), and if everything else about the agent and the external world is held fixed right up to t , then it seems just a matter of luck

whether the agent at t performs something that brings about A or not-A. It just so happened that the agent ended up in a possible world where she chose A, because she could just as well have ended up in a possible world where she chose not-A (Mele 2006). Since everything about the agent and the world is the same in both possible worlds right up until the moment of divergence, it seems unfair to blame the agent for choosing wrongly. Event-causal libertarians have given several different replies to this problem, which can be summarized as follows: although the agent's act is indeterministic, it is not random because whatever way the agent chooses, she chooses for her own reasons, consciously and rationally, whether her act is A or not-A (Kane 1996; Ekstrom 2000; Balaguer 2010). In this sense, she meets all reasonable compatibilist requirements, except, of course, the requirement for antecedent control, i.e., the requirement to be able to control her act deterministically. One might worry that indeterminism here is still a hindrance rather than a help, because it reduces control. However, event-causal libertarians can answer that this is the price to pay for freedom, and that libertarian agents have enough control to be responsible for their actions (Kane 1996, 144).

My objection does not challenge this answer. For the sake of argument, I grant that event-causal libertarians successfully reply to the luck problem by showing that neural indeterminism does not exclude moral responsibility. However, solving the luck problem will only demonstrate that neural indeterminism is compatible with moral responsibility but will not establish that neural indeterminism is necessary. Thus, the discrimination problem persists: in order to solve it, event-causal libertarians must go further; they additionally need to demonstrate why neural indeterminism is a prerequisite for moral responsibility.

One might think that Cristopher Franklin's (2018) version of event-causal libertarianism, which he calls "minimal libertarianism", offers a solution to this problem. Franklin asserts that indeterminism is morally relevant because it provides an agent with enhanced control. This is because while all of the agent's abilities necessary for free will are compatible with physical determinism, the opportunities to either exercise or not exercise these abilities at each particular time are not compatible with physical determinism. Thus, indeterminism at the moment of choice gives the agent the opportunity to act one way or another: the agent can choose to do A or not-A, given that everything else up until the moment of choice is held fixed. Consequently, this type of indeterminism provides enhanced control and is necessary for both free will and moral responsibility.¹⁶

But this type of enhanced control is still not the control over the agent's own neural states. In my thought experiments, the agents have alternative possibilities (they can form SFAs and L-free torn decisions). However, even if an agent like Rachel* performs her first L-free act by performing an indeterministic torn decision, the fact that her neural state was indeterministic was not under her control any more than the fact that Carol*'s neural state was deterministic was under Carol*'s control. We simply do not control how our neural states are realized. Thus, Franklin's minimal libertarianism doesn't solve the discrimination problem, either.

8. Conclusion

Event-causal libertarians postulate neural indeterminism in order to secure actualist alternative possibilities for agents

within a naturalistic worldview. They also claim that indeterminism is morally relevant: whether an agent was acting deterministically or indeterministically decides whether this agent is morally responsible for her act. I have considered three influential event-causal libertarian accounts: by Robert Kane, John Lemos and Mark Balaguer, and I have shown that in their current form, they do not provide an explanation why neural indeterminism is morally relevant. I have shown this by means of thought experiments where agents satisfy all of the requirements of the aforementioned libertarian accounts, except having neural indeterminism in the history of their choices. I claim that treating these agents differently from the agents who satisfy all the same criteria plus neural indeterminism is probably unfair. Unless persuasively argued for, this different treatment will constitute an unjust discrimination. So far, such an argument is not present in the debate. Christopher Franklin's (2018) minimal libertarianism does not provide a solution to this problem, either. Franklin explains that indeterminism at the moment of choice gives agents enhanced control by providing them opportunities to act one way or another, but neural indeterminism of the sort postulated by the event-causal libertarian accounts considered in the paper cannot enhance agential control.

I have also considered how the discrimination problem is related to the luck problem, but not identical to it. Luck problem is the problem of alternative possibilities given the same past: whether the agent can be responsible for their actions if the past is held fixed. I claim that even if the event-causal libertarian solution of the luck problem is successful, the discrimination problem remains. Solving the luck problem will only give us the conclusion that neural indeterminism is compatible with moral responsibility but will not prove that neural indeterminism is necessary. In order to solve the discrimination problem, event-causal libertarians must explain how neural indeterminism is relevant to our status of morally responsible agents.

Statements and Declarations

Conflict of interest

Competing interests: The author declares none.

Funding

This research was funded by DFG, Research Group Inductive Metaphysics (FOR 2495), Project A2.1.

Acknowledgments of personal gratitude

In memory of Robert Kane, a distinguished philosopher who kindly expressed interest in this project. His contributions to the development of event-causal libertarianism have been invaluable.

I would like to thank Karim Baraghith, Alexander Christian, Sergei Levin, Leonhard Menges, Jacob Rosenthal, Gerhard Schurz and Victoria Yugai for helpful comments on earlier versions of this paper.

Footnotes

¹ This technical incompatibilist requirement has emerged in the discussion of the consequence argument (Capes 2019; van Inwagen 1983). For compatibilist objections, see (anonimized; Campbell 1997). It is crucial to the libertarians that the alternative possibilities they require are incompatible with physical determinism, because otherwise their position becomes a version of compatibilism (Franklin 2018; Vihvelin 2013).

² In earlier works, Kane proposed longer definitions of UR (Kane 1996, 35). What matters for the purposes of this paper is that in all of Kane's definitions, UR is necessary for free will and incompatible with physical determinism.

³ CC is the best available set of conditions put forward by a compatibilist theory maintaining that physical indeterminism is compatible with free will but not necessary for it. Some authors call such a position 'supercompatibilism' (Vargas 2012), whereas other authors see this as a core compatibilist thesis since compatibilism is just the statement of the lack of incompatibility between free will and physical determinism (Vihvelin 2013). As a matter of fact, supercompatibilism is accepted by many contemporary compatibilists (Vargas 2012). Among the candidates for CC are compatibilist conditions discussed in (Campbell 1997; Jaster 2020; Sekatskaya & Schurz 2021; Vihvelin 2013). For the purposes of this paper, we do not need to choose a particular version of CC. It will suffice to assume that we choose the one which is the best among those which have been formulated so far, or, if no winner is available, is at least as good as the others.

⁴ According to Kane, whether an effort of will is truly indeterministic cannot be established introspectively by noticing some salient feature of its phenomenology (Kane 1996, 17). Introspection is not a reliable guide to judgments about what is going on when we experience an inner struggle characteristic of SFWs (Kane 2019, 151–153).

⁵ Dennett criticizes Kane's answer to his Luther example. He argues that publicly accessible facts about Luther's past and the whole history of his personal development are much more important than some hidden facts about the neural realization of his decision-making process (Dennett 2003, 129). This answer presupposes the importance of compatibilist conditions and downplays the importance of libertarian conditions for free will.

But since Kane accepts that meeting compatibilist conditions is important (1996, 15, 32) – indeed, as he clarifies in his recent work, even necessary for meeting libertarian conditions for free will (2019, 149) – Kane can argue that although the circumstances of Luther's biography do matter, the facts about his SFAs matter as well. Although they might not be discoverable by the usual means of observation or introspection, they are still of the utmost importance because they require AP and guarantee UR. Presumably, this is why Kane's later views have not changed regarding the Luther example in particular and the importance of SFAs in general (Kane 2007a, 2011, 2019).

⁶ I would like to thank an anonymous reviewer of an earlier version of this paper for making this point.

⁷ As I argued in Sections 2 and 3, UR is necessary for all cases of moral responsibility, either directly when an agent is ultimately responsible for some SFA, or derivatively when an agent is morally responsible for an action that is a deterministic consequence of an SFA for which she was ultimately responsible. HM, as I have shown, is also necessary for moral responsibility because it establishes a criterion of justified derivation of moral responsibility for an action

deterministically following from one's character formed by SFAs. In (1996, ch. 6), Kane explains that moral responsibility must be understood in an absolute sense, not relativistic or consequentialist, and that just punishment must be based on one's moral responsibility for an illegal act, not on relativistic or consequentialist considerations.

⁸ Recent work in neuroscience demonstrates that retrospective predictions of some narrowly defined class of actions, such as button-pressing or wrist-flexing, can already be done with a high degree of probability for human subjects (Soon et al. 2008). Whether this data gives us a reason to doubt the existence of free will is a hotly contested issue (Haggard 2019). We do not need to take any position on this issue here, because all that we need for the argument to go through is the conceivability of this thought experiment.

⁹ This intuition is never defended or questioned, but simply postulated as immediately obvious (cf. Kane 1996; 2007a; 2019). Its relation to other aspects of Kane's account were explored in Sections 2 and 3 of this paper.

¹⁰ For the role of intuitions and the burden of proof strategy in the free will debates, see (Jackson 2008; Lycan 2003, Nahmias et al. 2005).

¹¹ My general strategy is showing that this clash of intuitions is a case of one man's *modus ponens* being another man's *modus tollens*. Namely, what is more plausible: that our intuitions about free will require acceptance of HM, or that C-people should not be discriminated against if they do not differ from L-people in anything except HM-related properties? This helps to address the following objection (I thank [redacted] for raising it). If Kane is right, moral responsibility requires performing *N* SFAs. Consequently, it would be unjust to punish those who do not satisfy HM, and it would be just to punish those who do. Here is how one could argue for it: "1. The fact that L-people performed *N* SFAs partly explains why they have free will. 2. That people (knowingly) do wrong out of free will makes it the case that they deserve blame for it. Therefore, they deserve blame". However, the discrimination objection does not question the internal connection between free will, SFAs and moral responsibility within Kane's theory. It raises a conflicting intuition of injustice: why is the historical fact that a person has performed *N* SFAs relevant to this person's moral status, if this fact is neither psychologically accessible to this person (as L-people and C-people can be psychologically indistinguishable) nor does it give an L-person any additional abilities (as L-people and C-people equally satisfy AP: they are able to act one way or another given the same past by performing an SFA)? If the proponent of HM replies: it is relevant because it follows from the UR-intuition, the opponent can say: why should I prefer your UR-intuition to my unfair discrimination intuition? The answer to this cannot be: because the UR-intuition is clearly right! Which one of them is right is precisely the point of dispute.

¹² I would like to thank an anonymous reviewer of an earlier version of this paper for making this point.

¹³ The reasons for the replacement are complicated, and given in (Balaguer 2010, ch. 2), but can, I think, be summarized as follows: one of the problems in the discussions of free will consists in trying to establish what the term 'free will' means in ordinary language, or what it should mean if the ordinary language term were clarified by conceptual analysis and/or new information from the science of human psychology. This conceptual question has generated many discussions, but they are not relevant to the following 'metaphysical' question: do people actually have the ability (or the family of abilities) that libertarians associate with free will? According to Balaguer, it is obvious that people have the abilities that

compatibilists associate with free will (Balaguer 2010, 54). Therefore, the question whether people have the abilities that compatibilists call ‘free will’ is not metaphysically interesting.

¹⁴ A similar principle has been invoked by David Widerker (2005), who argued that blame is morally justified only if there is an informative answer to the blamed person’s question “What should I have done instead?” In other words, moral blame presupposes that a person had alternative courses of action which she could have chosen.

¹⁵ This vague formulation reflects the fact that different versions of event-causal libertarianism disagree about the precise location of indeterminism in the decision-making process. All event-causal libertarians agree that there must be indeterminism somewhere in the decision-making process, but where and when, exactly? Some alternatives are the following: indeterminism should occur at the moment of different options coming to one’s mind (Mele 2006), or at the moment of deliberation and weighing of different options already present in one’s mind (Kane 1996; Ekstrom 2000), or precisely at the moment of making the choice to act one way or another (Balaguer 2010; Franklin 2018).

¹⁶ For another criticism of this position, different from my discrimination objection, see Law (2022).

References

- Balaguer, M. 2010. *Free Will as an Open Scientific Problem*. MIT Press.
- ———. 2019. “Free Will, Determinism, and Epiphenomenalism.” *Frontiers in Psychology* 9:2623.
- Campbell, J. 1997. “A Compatibilist Theory of Alternative Possibilities.” *Philosophical Studies* 88 (3): 319–30.
- Capes, J. 2019. “What the Consequence Argument Is an Argument For.” *Thought: A Journal of Philosophy* 8 (1): 50–56.
- Dennett, D. 1984. *Elbow Room: The Varieties of Free Will Worth Wanting*. MIT Press.
- ———. 2003. *Freedom Evolves*. New York: Viking.
- Ekstrom, L. 2000. *Free Will: A Philosophical Study*. Focus Series. Boulder, Colo: Westview Press.
- Franklin, C. 2018. *A Minimal Libertarianism: Free Will and the Promise of Reduction*. Oxford University Press.
- Haggard, P. 2019. “The Neurocognitive Bases of Human Volition.” *Annual Review of Psychology* 70 (1): 9–28.
- Jackson, F. 2008. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Reprinted. Oxford: Clarendon.
- Jaster, R. 2020. *Agents’ Abilities*. De Gruyter.
- Law, A. 2022. “What Does Indeterminism Offer to Agency?” *Australasian Journal of Philosophy* 100 (2): 371–85.
- Lycan, W. 2003. “Free Will and the Burden of Proof.” *Royal Institute of Philosophy Supplement* 53: 107–22.
- Kane, R. 1996. *The Significance of Free Will*. Oxford University Press.
- ———. 2007a. “Libertarianism.” In *Four Views on Free Will*, by John Martin Fischer, Robert Kane, Derk Pereboom, and Manuel Vargas. Oxford: Blackwell Pub.
- ———. 2007b. “Response to Fischer, Pereboom, and Vargas.” In *Four Views on Free Will*, 166–83. Malden, MA: Blackwell.
- ———. 2011. “Rethinking Free Will: New Perspectives on an Ancient Problem.” In *The Oxford Handbook of Free Will*,

381–404. Oxford University Press.

- ———. 2019. “The Complex Tapestry of Free Will: Striving Will, Indeterminism and Volitional Streams.” *Synthese* 196 (1): 145–60.
- Lemos, J. 2018. *A Pragmatic Approach to Libertarian Free Will*. Routledge/Taylor & Francis Group.
- Mele, A. 2006. *Free Will and Luck*. Oxford University Press.
- Nahmias, E., Morris, S., Nadelhoffer, T., and J. Turner 2005. «Surveying Freedom: Folk Intuitions about Free Will and Moral Responsibility». *Philosophical Psychology* 18 (5): 561–84.
- Sekatskaya, M. and Schurz, G. 2021. “Alternative Possibilities and the Meaning of ‘Can’.” *Dialectica* 75(3).
- Soon, C., M. Brass, H.-J. Heinze, and J.-D. Haynes. 2008. “Unconscious Determinants of Free Decisions in the Human Brain.” *Nature Neuroscience* 11 (5): 543–45.
- Van Inwagen, P. 1983. *An Essay on Free Will*. Clarendon Press.
- Vargas, M. 2012. “Why the Luck Problem Isn’t.” *Philosophical Issues* 22 (1): 419–36.
- Vihvelin, K. 2013. *Causes, Laws, and Free Will: Why Determinism Doesn’t Matter*. Oxford University Press.
- Widerker, D. 2005. “Blameworthiness, Non-Robust Alternatives, and the Principle of Alternative Expectations.” *Midwest Studies in Philosophy* 29 (1): 292–306.