Reasons-Responsive Theories and the Nature of Reasons

Qiuxuan Zheng[1,*]

[1] *Xi'an Jiaotong-liverpool University, China*
[*]*Email: Qiuxuan.Zheng18@student.xjtlu.edu.cn*

ABSTRACT

In the contemporary free will debate between compatibilism — the thesis that free will is compatible with determinism, and incompatibilism — that free will is incompatible with determinism, many scholars are sympathetic to compatibilism, yet disagree in how the position is best characterized. As one of the most important branches of *source compatibilism*, reasons-responsive theories attempt to address an important issue, namely, what type of person can be regarded as moral agents. There are several controversial points within the theory. This paper aims to show that reasons-responsive theories fail as a solution to the free will debate, to the extent that they remain silent on the issue of the nature of reasons. It is this silence and inadequacy that constitute a rebuttal to reasons-responsive theories. In what follows I will propose a dilemma to reasons-responsive theories: if moral reasons are objective in the sense that one could be wrong about them, it appears that one must be a moral expert to possess the so-called "reasons-reactivity", a necessary condition for the possession of free will. This will later be shown to be deeply implausible. If moral reasons are instead subjective in the sense that whether one has a moral reason to do X crucially depends on his mental states, then reasons-responsiveness seems to have no connection with free will at all. In addition, there are independent grounds on which the subjective account of the nature of reason can be challenged. Therefore, no matter which road reasons-responsive theorists choose to go down, there are serious objections that cast doubt on their plausibility. At the end of the paper, I will touch on the question of why reasons-responsive theories are doomed to failure due to their lack of concern with the agent's own psychology.

Keywords: Free will, Determinism, Reasons-responsive theories, Source compatibilism.

1. INTRODUCTION

Prior to contemporary compatibilism, classical compatibilists believe that an agent could be considered free as long as their behaviour is not restricted by external conditions. However, contemporary compatibilism rejects this claim and dismisses the restriction on the agent's psychological structure, believing that the agent will not possess free will until they are unconstrained on both the external and the psychological level. Therefore, the factors implicated in the agent bringing about their action must be a motivation under their control. However, it remains unclear how to determine if a factor is controlled by an agent. The notion of reasons-responsiveness affords an appealing new direction for solving this question.

Firstly, this theory denies the key premise of classical compatibilities, in that it takes freedom, for example, the control condition for moral responsibility, to be independent of the ability to do otherwise. Moreover, it suggests a source-compatibilist-friendly view, according to which only *source freedom* (viz. the agent is the source of their action) is required by moral responsibility. This theory has roots in the Aristotelian ethical tradition which believes only human beings as rational animals are aware of how to make informed decisions and thereby be held morally responsible for their actions. On this basis, reasons-responsive theories aim to distinguish the unimpaired functioning of normal human agency from other agents, including young children and agents acting on compulsive mental illness, such as phobias.

This is further used for distinguishing both and is the central thesis of reasons-responsive theory, which is the *actual-sequence analysis of using*. However, there are two caveats to this term. The first is *guidance control*, which means the mechanism that issues in the action is his own reasons-responsive mechanism, according to

Fischer and Ravizza [1]. The second is *actual-sequence*, meaning whether an agent is morally responsible for his actions relates only to what they have actually done in reality, rather than the situation in the counterfactual conditional.

This viewpoint is further illustrated by *Frankfurt's case* [2]. In this case, the agent still acts with guidance control, due to the mechanism of his actions being responsive to the relevant reasons. Therefore, the agent is reasons-responsive only due to a reasons-responsive mechanism.

1.1. Fisher and Ravizza's theory

Reasons-responsive theories fall into two broad categories — mechanism-based theories and agent-based theories. The mechanism-based theories narrow down the spectrum of causation and focus solely on what an agent has done in an actual scenario and the reasons for their behaviour, considers the causal role the agent's deliberative characteristics played in the actual sequence of events that led to their behaviour and calls it the mechanism that produces their actions.

Fischer and Ravizza emphasise that the extent to which agents are responsive to reasons is different. Specifically, agents who are *strongly* reasons-responsive respond in different ways only if they have sufficient reason to do so, whereas agents who are *weakly* reasons-responsive are prone to act differently as long as there is a reason. However, the former excludes responsibility for many unethical and weak-willed acts, whereas the criteria of the latter are too weak.

Therefore, Fischer and Ravizza propose two key elements for ensuring that the theory becomes a *moderate* reasons-responsive theory: a receptivity component and a reactivity component. These refer to the scope of reasons and the degree of responsiveness to the reasons. Receptivity is established in order to distinguish between individuals who are mentally disturbed, young children and animals, and normal people who are capable of using free will. This ensures that the agent is rationally responsive to the reason. Reasons-reactivity implies that agents react differently based on different types of rationality. Fischer and Ravizza noted that reactivity is *all of a piece* [1]. That is, a mechanism's reacting to a sufficient reason to do otherwise in some possible world establishes that mechanisms of this type have the general capacity to react differently to any sufficient reason to do otherwise.

1.1.1 Mechanism

However, Fisher and Ravizza's mechanism-based theory has been challenged with a number of criticisms. The main criticism is that it is not even clear how the mechanism is individuated. To illustrate this with an example, addicts normally take drugs because they are addicted to them, but if they are aware that they will die immediately afterward, they will not take drugs again.

Opponents note that the underlying mechanisms that drive the two situations are importantly different. . In this case, the addict's desire not to die has played a significant role in his deliberating process, such that the second case incorporates a new mechanism [2]. In addition, Watson notes that this example demonstrates reactivity is not a monolithic entity, since the agent responding to a particularly strong reason does not mean they will respond to all reasons [3]. Therefore, reactivity to a restricted set of reasons does not suffice to establish full-fledged reasons-responsiveness.

1.1.2 Ownership

Another powerful criticism highlights that the ownership criteria are rather obscure. In Fischer and Ravizza's specification of moderate reasons-responsiveness, the concept of ownership is used to ensure that the mechanism which produces an agent's actions is his own [1].

They note that ownership must satisfy three conditions: "First, the agent must come to view herself, when acting from relevant mechanisms, as an agent, capable of shaping the world by her choices and actions. Second, she must see herself as an apt target of others' moral expectations and demands as revealed in the reactive attitudes. Third, the beliefs satisfying the first two conditions must be based, in an appropriate way, on the individual's evidence." [2]

However, it is extremely difficult to determine the exact meaning of ownership from the definition. This ambiguity can also be seen in examples. Watson argues against mechanism-based theories with the example of Goldie in the voting booth. He suggested that when Frankfurt interfered with Goldie's vote, he had no ownership of the decision as it was Frankfurt's interference, rather than Goldie's own free will, which led to the final vote result [3]. However, the Frankfurt interference in Goldie's vote does not appear to be fundamentally different from any other interference which might occur in daily life. As Watson stated, perhaps a nutrition bar would be enough to make Goldie behave differently than before. If ownership can be deprived so easily, it appears that it is impossible for a mechanism to be an agent's own at all. Therefore, the concept of mechanism needs serious refinement.

1.2 Agent-based theory

Although mechanism-based reasons-responsive theories are subject to the above objections, there are compelling reasons for adopting agent-based reasons-responsive theories, which are arguably immune to

these objections. This is mainly due to the fact that agent-based theories do not appeal to the notion of mechanism at all to vindicate their version of the reasons-responsive theory.

The most articulate agent-based theories were developed by McKenna and Sartorio. They use Frankfurt's case to provide evidence illustrating how an agent is responsive to a variety of reasons, particularly how an agent reactive to these reasons has source freedom. Initially, it appears that agent-based theories do not fit well with Frankfurt's diagnosis of the standard cases: how could Jones be free regarding his action of killing Smith when Black's presence guaranteed that Jones could not have been responsive to the reasons to not kill Smith?

McKenna's solution stems from a subtle revision of the counterfactuals. While it is false that <if reason R had been present, Jones would have refrained from killing Smith>, the following counterfactual is nevertheless true: <If R had been present and if Black had not activated the interfering device, Jones would have refrained from killing Smith>. In addition, McKenna notes that the same effect can be achieved through weakening the consequent, suggesting that "what is required is not the capacity to do otherwise in the presence of sufficient reasons to do otherwise but, for example, the capacity to fail to act on the basis of one's own reasons" [4]. In Frankfurt's case, this means if Jones had been aware of reason R (for example, Smith's children were with him), then Jones would not kill him for his own reasons.

Sartorio is another preeminent scholar who advocates agent-based theory. She proposes a more elegant version of the reasons-responsive theory that avoids appeal to the notion of mechanism altogether; that is, an elegant way of explaining the agent's moral responsibility by appealing to the actual causes of her action [5]. This is because, in Sartorio's view, the absence of certain reasons may be part of an agent's actual reason for acting the way they behave [6]. She uses Frankfurt's case to further validate her viewpoint. In the case above, Jones would not kill Smith if he had been aware that his children were with him. However, in actual-sequence, Jones killed Smith precisely because that reason, which Jones would have responded to if present, was absent.

2. THE NATURE OF REASONS

Agent-based reasons-responsive theories have succeeded in avoiding the limitations of mechanism, but in what follows I will show that other objections are targeting reasons-responsive theories in general. Note that many reasons-responsive theorists are silent about the nature of (moral) reasons. Sartorio, for example, said that scholars can be neutral and do not need to take a specific stance regarding what reasons are in order to endorse a reasons-sensitivity view of freedom. In response to this line of silence, I argue that what reasons are affecting whether or not the reasons-responsive theories are tenable. In the next paragraph, I will propose an argument, which takes the form of a dilemma. It shows that moral reasons are either objective or subjective, but either way can raise serious problems for reasons-responsive theories in general.

2.1 Objective

Reasons are objective, which means it is an objective matter of fact that <if reason R is present, then the agent should do A>. For example, <if the room is on fire, then people should escape>, it appears to be a paradigmatic objective reason. This fact is independent of human attitudes. Therefore, if the relationship is objective, then knowledge of this reason is no different from knowledge of any other objective facts in the world (e.g. <1+1=2>, or <the earth orbits around the sun>). Therefore, an agent may fail to recognise a reason, or they may mistake a non-reason as a reason. And when they do so, they have made a cognitive mistake. For example, an agent who has a neurological disorder may run from a room as they believe that it is on fire (when it is not). However, observers would intuitively say that there was *no reason* for them to run out of the room, since the reason a disordered person took was not actually a reason. Therefore, a rational person is who is able to determine whether something is a real reason.

The objectivity of reasons is implicitly presupposed in Fischer and Ravizza's theory. They believe that one must correctly recognize moral reasons to have free will. For example, when a condition of moral competence is stipulated, they note that "the kind of responsiveness required for moral responsibility ought to be characterized not merely as a responsiveness to reason, but rather as a responsiveness to a range of reasons that include moral reasons". More importantly, the agent cannot be insane in the sense that they are only responsive to bizarre reasons.

Reason-receptivity adds to weak reactivity the requirement that a person's responsiveness exhibits "an understandable pattern of reasons-recognition, minimally grounded in reality" [7]. Watson provides an example of Brown who only buys baseball tickets if they cost exactly $1,000. He will not buy them even if the ticket price is less than $1,000. In this case, it is fair to say that Brown has reason-receptivity to a certain extent. However, according to Fischer and Ravizza, this condition is insufficient as <buying baseball tickets only if they cost exactly $1,000>, which is a bizarre reason. This demonstrates that they have a clear objective conception of what the reasons are. They believe that if a person possesses reason-receptivity in a real sense,

then she should not be responsive only to bizarre reasons.

2.1.1 Moral BIV

In addition, a more pressing issue is that, if the nature of reason is objective (allowing the happening of cognitive error), then according to reasons-responsive theorists, it would be impossible for one to be systematically wrong about reasons for actions yet still have free will. That is to say, agents who do not recognize some objective, mind-independent facts cannot have free will. Now I describe a counterexample to this claim, which calls Moral BIV. The Moral BIV was born out of the original brain-in-a-vat scenario, where the agent holds beliefs about her immediate environment just like other people, and even though her beliefs are systematically false, she is justified in believing these things to precisely the same extent as other people [8].

Similarly, a Moral BIV holds moral beliefs just as ordinary people do; her belief may even be justified. In addition, similar to the original BIV's beliefs being systematically false, the Moral BIV's moral beliefs are systematically false as well (e.g. she believes that killing infants is morally permissible, that lying is morally praiseworthy in all cases). That is, the Moral BIV is systematically wrong about moral reasons for actions, and therefore fails to be responsive to moral reasons. However, it should be obvious to anyone that even a Moral BIV can have free will!

Now, reasons-responsive theorists will certainly object that moral BIV are not morally competent and cannot bear moral responsibility. On their view, given that a moral BIV is entirely wrong with respect to what are moral reasons to do, she does not qualify as a moral agent and therefore does not possess free will. In other words, a moral BIV does not meet the conditions necessary to possess free will. A moral BIV is similar to a brainwashed person whose actions are not the result of rational cognition and deliberation but simply what the person who brainwashed them wants them to do. In this case, they would argue that it is intuitively clear they are not morally responsible for wrongful actions.

This objection confuses possession with use. To clarify this point, a person possesses reason-receptivity in a real sense, then they should not be responsive only to bizarre reasons.

However, the problem is that on this view, one must be a moral expert in order to have free will and morally responsible agency. This is deeply implausible for various reasons. To begin with, a significant number of moral judgments are moot points. For example, there have been debates over whether or not a woman should be allowed to have an abortion. Many people believe that abortion is murder, basing their belief on the fact that a foetus is a life. Others concede that although a foetus is indeed a life, it should not be taken as a reason to oppose abortion. Instead, they claim that the pregnant should have control over her own body (more specifically, her uterus) and have the right to choose abortion. This controversy is still a moot point, and it remains unclear who the experts are. But if it is an objective matter of fact whether <a foetus is a life> constitutes a moral reason to oppose abortion, then those who happen to be wrong turn out to be unresponsive to reasons on this particular matter!

The moral BIV will be compared to the original BIV. For the original BIV, the sensory signal does not come from the outside world, but from the supercomputer. Although their beliefs are false, it does not prevent them from having the ability to form beliefs. For the original BIV, possession is the ability to form cognition based on visual signals, and use refers to the ability to act under the guidance of cognition. Although the original BIV fail to use his beliefs, he is still able to process his beliefs, which ensure he to act like a believer.

Similarly, to the moral BIV, possession is the cognitive capacity to distinguish right from wrong, and use asks for the volitional capacities to conform one's conduct to that normative knowledge. It appears that although a moral BIV mistakes non-reasons as reasons, it does not deprive the moral BIV of his agential power. The moral BIV is simply unlucky in the sense that he is not in the position to properly use his reasoning capacity to derive correct moral judgments. In other words, there is a difference between understanding a moral norm and embracing it. For example, supposing <if Smith has five children, he should not be murdered> is a moral norm, a moral BIV will not embrace this norm. However, they understand this norm perfectly. This is precisely due to their ability to consider the reason that a moral norm is wrong and they can have the exact opposite thought and use it to guide their behaviour. In other words, a moral BIV is capable of understanding the moral norm, but freely and voluntarily refuses to embrace it. As long as the moral BIV understands the norm, he still has the free will required to bear moral responsibility.

Another objection by reasons-responsive theorists may be that the moral BIV is after all not morally responsible for his actions, since he does not meet the relevant cognitive condition. My reply is that, although a moral BIV cannot be morally responsible for his action, it is not because she does not have free will. I am going to argue that embracing a moral norm is not a necessary condition for possessing free will. This is because, to be morally responsible, an agent needs to satisfy both certain control conditions (he must have the ability to control his actions, the so-called ability to choose freely etc.) and certain cognitive conditions (he must know what he is doing etc.). Free will only involves the control condition of moral responsibility,

but not the cognitive condition. As a result, this objection still cannot prove that moral BIV does not have free will. While the reasons-responsive theories must prove that moral BIV does not have free will, because these theories establish the distinction between people with free will and people without free will. If even a moral BIV could have free will, these theories would fail.

2.2 Subjective

Now I am turning to the second horn of the dilemma, namely, reasons such as <if reason R is present, then the agent will do A> are subjective. In this situation, "subjective" means that if an agent takes R as a reason to do A, then R is ipso facto a reason to do A. In this context, R is a reason to do A does not imply that R is a good reason to do A. If this relationship is subjective, the first objection is the false reason problem. For example, if <my house is on fire> is a reason for me to leave the house, then there is a reason for me to leave the house. However, if I am mistaken and my house is actually not on fire, then it is natural to assert that there is no reason for me to leave the house. But if this relationship is subjective and I take <my house is on fire> as a reason to leave the house, <my house is on fire> as a reason for me to leave the house. This constitutes a contradiction.

Furthermore, if this relationship is subjective, there can be reasons to do anything. In this context, reasons-responsiveness would come too cheap. Because even an insane is able to responsiveness to reasons. Initially, this theory was created to capture the morally important difference between people who are rational and people who are not. However, if this relationship is subjective, then an insane person can be perfectly reasons-responsive. Recall Watson's Brown example (Brown will not buy a ticket until it is exactly $1000), it is also unclear why it is positive to be reasons-responsive.

In addition, a more serious problem is about moral responsibility. That is, if a reason is subjective, even if an agent cannot bear moral responsibility, they can still possess free will. For example, if an agent unknowingly kills a baby by throwing away a package (containing a baby), they are clearly not morally responsible for the act. Conversely, if this person grew up as a moral BIV and believed killing to be good, they would throw away the parcel even though they knew that the baby would die if they did so. However, intuitively, they should not take moral responsibility for this (it is not their fault they did it, they were just brought up with the wrong morals).

Based on the previous argument, it is apparent that they still possess free will. This problem exposes another flaw in the theory that the reasons-responsive theory is based on the misplaced basis of moral responsibility. From a moral responsibility perspective, if an agent can undertake moral responsibility, the certain control conditions (they should have the ability to control their behaviour and have so-called free choice, etc.) and knowledge conditions (they should understand what they are doing, etc.) must be satisfied. As the reasons-responsive theory appears to be classified as a knowledge condition rather than a control condition, it is impossible to fully describe free will in a theoretical sense. Returning to the previous example, even if we believe that a moral BIV cannot bear moral responsibility, the reason appears to be that they do not satisfy the knowledge condition (they do not know what type of behaviour is morally correct), rather than they do not possess free will. In this way, the reasons-responsive theory neither has a connection with free will, nor can it be used for resolving compatibilism versus incompatibilism debate.

3. CONCLUSION

I conclude this paper with another, more general objection to reasons-responsive theories, and argue that mesh theory (another compatibilist approach that situates free will within the agent's own psychological architecture). They fail to provide an explanation of why free will and moral responsibility are linked to reasons-responsiveness in the first place. Neither reasons-responsiveness can be recovered from the agent's psychological architecture, in which case reasons-responsive theory is inadequate; nor it can be recovered, in which case it appears superfluous, as we can skip this theory and directly discuss psychology and freedom. This gives mesh theory advantages over reasons-responsive theory. Mesh theory observes the psychology of agents and believes whether a person is free is ultimately determined by their psychology. In this way, mesh theory attempts to provide the ultimate answer to what type of psychology produces free will, as a result, it has an advantage over reasons-responsive theory in terms of explanatory power. Furthermore, when moral responsibility is discussed, the focus is not on being morally responsible for something (viz. people can be held morally responsible for a certain action) but on the possession of morally responsible agency (viz. what type of people have the ability to be morally responsible agents). In other words, there is great interest in what makes a person considered to be a morally responsible agent. It appears that an explanation would better lie within the agent's own psychology. This is further evidence that shows why mesh theory is more worthy of pursuit.

AUTHORS' CONTRIBUTIONS

Qiuxuan Zheng is the only author of the paper, who put forward the ideas, and completes the writing and modification of the paper.

REFERENCES

[1] Fischer, J. M., Ravizza, M. (1998) *Responsibility and Control: An Essay on Moral Responsibility.* Cambridge: Cambridge University Press.

[2] McKenna, M., Pereboom, D. (2016) *Free Will: A Contemporary Introduction.* New York: Routledge press.

[3] Watson, G. (2001) 'Reasons and Responsibility(*)', *Ethics*, 1 January. Available at: https://search.ebscohost.com/login.aspx?direct=true&db=edsgea&AN=edsgcl.72052226&site=eds-live&scope=site

[4] McKenna, M. (2013) *Reasons-Responsiveness*, Agents and Mechanisms. Oxford: Oxford University Press.

[5] Sartorio, C. (2016) '*Causation and Free Will'*. Oxford: Oxford University Press.

[6] Stanford Encyclopedia of Philosophy (2018) Available at: https://plato.stanford.edu/entries/freewill/#ArguForRealFreeWill

[7] Wolf, S. (1990) *Freedom within Reason*. New York: Oxford University Press

[8] Lehrer, Keith., Cohen, S. (1983) 'Justification, Truth, and Coherence', *Synthese*, 55(2), pp. 191-207. Available at: https://search.ebscohost.com/login.aspx?direct=true&db=edsjsr&AN=edsjsr.20115868&site=eds-live&scope=sit