

Dominika Palka

University of Warsaw

Reasons-Responsiveness: One More Look at the Agent-Based and Mechanism-Based Accounts

According to Fischer and Ravizza, having alternative possibilities of action is not a necessary condition of moral responsibility. The authors are convinced by the Frankfurt examples and motivated to hold this position due to the unresolved issue of compatibility of determinism and the possibility to act otherwise. Instead, they propose that what we need for holding people morally responsible is that they possess a kind of control called “guidance control”; and this is dependent on the mechanism which leads to action’s being moderately reason-responsive. The notion of moderate reasons-responsiveness (MRR) is explicated as follows: a type of mechanism M is MRR if and only if (1) in many possible worlds where it operates and S has a sufficient reason to do otherwise, S recognizes that reason as sufficient; (2) S’s disposition to recognize different reasons forms an intelligible pattern; (3) in at least one of those worlds, S acts on the recognized sufficient reason. After the authors present their account in their book *Responsibility and Control. A Theory of Moral Responsibility* (1998), a lot of discussion and criticism emerged. One of the main points of criticism was the vagueness of the notion of mechanism and the lack of identity criteria for different types of mechanisms. This led some philosophers, Carl Ginet and Michael McKenna among them, to adopt an agent-based reasons-responsiveness theory. This modification faces other problems. First, it has been suggested that it is vulnerable to Frankfurt-like examples, since the agent with whom the counterfactual intervener is connected is not reasons-responsive, although she is morally responsible. This charge is rejected both by Ginet and McKenna. However, the latter admits that there are possible examples, concerned with “intervener” who is internal to the agent, that may threaten the agent-based theory. One of his undeveloped examples, constituting the second problem, is the situation in which the mentally ill agent acts while his sickness is not active, but it would activate and prevent him from acting otherwise in alternative scenarios. In this paper, I present examples based on McKenna’s suggestion and argue that they pose a much more serious problem for agent-based account than McKenna would like to admit. I also present a possible counterexample to both agent-based and mechanisms-based account concerning a recovered addict who refuses to take a drug, yet seems to be morally responsible (in this case – praiseworthy) for her resistance and yet is not reasons-responsive in the right way.

Keywords: incompatibilism, semi-compatibilism, reasons-responsiveness, moral responsibility

1 Why Reasons-Responsiveness Theory?

In 1998 John Martin Fischer and Mark Ravizza proposed, in their insightful and heavily discussed book *Responsibility and Control. A Theory of Moral Responsibility*, the new approach to assign moral responsibility to acting agents. They developed this particular view because of two motivations, which can be referred to as positive and negative incentives. First – and this is the negative incentive – the authors acknowledge that there is a potential threat from physical determinism to freedom understood as a possibility to act otherwise. Although neither of them claim to be fully convinced by incompatibilist arguments, they admit that a version of consequence argument proposed by Peter van Inwagen (van Inwagen, 1975: 188–194) is strong enough to urge us to look for accounts of responsibility that would allow us to hold people morally responsible even if both determinism and incompatibilism appear to be true.

The second, positive incentive, is the appeal of Frankfurt examples (Frankfurt, 1969: 835–836) concerning the possibility of being morally responsible for acting in a particular way in the situations where one has no alternative possibilities of action. Those are the cases of the following structure:

1. An agent S is driven to a decision D by her values, beliefs, desires /.../; and then she performs the action A as a consequence of that decision;
2. There exists a factor (originally: another agent) that would intervene if S was about to make any decision other than D in such a way that A would anyway make decision D and consequently do A;
3. It is intuitively plausible that S is morally responsible for A.

Convinced by these type of examples, Fischer and Ravizza try to build the theory that accounts for the responsibility of the agent in situations that fall under the presented scheme.

2 The theory – some important features

The most important notion of Fischer and Ravizza's theory and other theories that are based on it yet differ in some details is the notion of reasons-responsiveness. In the most general manner, this notion can be introduced as follows:

‘An entity E is reasons – responsive if and only if in an appropriate range of scenarios E recognizes reasons for acting differently than it acted and it modifies its behaviour in accordance with those reasons.’

Simple as that, but there is some nuance to be noticed here: to say that something is reasons-responsive in the specified sense is not to say that ‘it could have modified its behaviour (in the actual scenario); although it is possible to either argue that this is a real meaning of a phrase or to make a revisionist twist and claim that it should mean exactly that.

It is useful to inquire into the intuitions that underlie the view that possessing a property at which such a notion points at least partially grants or grounds moral responsibility – or is a necessary condition of it, dependent on a version of the theory. In a wide sense it focuses on the rationality of the entity whose responsibility is being assessed. In the case of an agent, this demand for rationality stands for the fact that it is generally agreed that agents who we do not consider as sane in a relevant sense should not be held responsible. This sanity can be roughly described as consisting of such traits as:

- having beliefs that are at least partially grounded in reality;
- performing thought and decision processes that are at least partially understandable;
- displaying at least a partial consistency between one’s beliefs, desires, and actions.

What exactly stands for the “entity” in the presented definitional scheme? There are two options, concerning which there are huge disagreements among philosophers who adopt reasons-responsiveness theory. Fischer and Ravizza insist on treating mechanisms that lead to action as subjects of reasons-responsivity. It is so because, so they claim, if we demand that the agent herself is a subject of reasons-responsivity, the account falls prey to Frankfurt-like examples. Recall that, in Frankfurt examples, there were no scenarios in which the agent acted otherwise: in every scenario in which she was about to decide to act differently, the intervener intervened. In a sense, it was something *about the agent* that there has been an intervener associated with her. On the other hand, it was not something in any way about the particular mechanism that led to action. Conversely, the key feature of this mechanism – the actual scenario’s mechanism – is that it is the agent’s own mechanism that operates without any external influence. Holding the mechanism fixed with this feature while considering different scenarios allows us to explain why, when we consider some particular situation, even with an assumed counterfactual intervener, we can focus on what the agent would have done *if no intervener were to be present* when assessing her moral responsibility.

However, the mechanism-based approach faces a serious problem: how to individuate between mechanisms operating in different scenarios? How – and if – can we (not too vaguely) specify what is the mechanism that operates in the actual scenario? Fischer and Ravizza themselves admit that they are not in a position to provide any good specification of this notion and that they are unable

to provide strict criteria for the individuation of mechanisms; neither was there anyone who would develop their account in a way that provides these missing elements. Alternatively, agent-based variations of the theory were developed to avoid the problem posed by mechanisms.

3 Into details

Since agent-based approaches are not completely independent types of reasons-responsiveness theory, but rather its modifications based on a common core that is provided by Fischer and Ravizza's analysis of the notion of reasons-responsivity, I will now use some space to recall their detailed explication of this notion. Until now, I was using the words "reasons-responsiveness" without further qualification. From now on, I will be talking about *moderate* reasons-responsiveness (MRR). Reasons-responsiveness comes in degrees, and MRR stands for exactly how responsive the agent must be to be possibly regarded as a proper subject of moral responsibility. The best definition of MRR, extracted carefully from Fischer and Ravizza's book, is provided by Carl Ginet in the following way:

Moderate reasons-responsiveness for a positive action

The kind of mechanism that issued in S's doing A was moderately reasons-responsive if and only if:

(1) in many different possible worlds where the same kind of mechanism operates in S and S has a different sufficient reason to do otherwise, including some where the reason is a moral one, S recognizes that she has a sufficient reason to do otherwise;

and

(2) S's dispositions to recognize various different reasons (including some moral ones) for doing otherwise form a pattern that is rationally intelligible (given S's beliefs and basic values)

and

(3) in one of these possible worlds S recognizes the sufficient reason to do otherwise and chooses to act, and acts, otherwise, for that reason. (Ginet, 2005: 232–233)

Fischer and Ravizza understand actions as bodily movements. They must therefore provide the extension of their account to the responsibility for consequences of actions, where consequences are taken to be states of affairs. Thus, the consequence of my action A, say, raising my hand, is a state of affairs: *that my hand is raised*; and also some further consequences, for example: *"that I voted for obligatory vaccinations"* or any other states of affairs that results from my bodily movement.

voted for obligatory vaccinations” or any other states of affairs that results from my bodily movement.

The formal account of the consequences of actions consists in two parts: (1) the requirement that the mechanism that led to a bodily movement is reasons-responsive, in a way that was defined in the preceding definition; and (2) the requirement that the process leading to a consequence of an action is properly sensitive to the bodily movement. Formally:

If S were to move his body in way B* [different than what she actually did at T] at T, all other triggering events (apart from B*) which do not actually occur between T and T+i or which actually occur and bring about C simultaneously or subsequently to T+i were not to occur, and a P-type process were to occur, then C would not occur. (Fischer and Ravizza, 1988: 120–121)

Equipped with these definitions, we can proceed to the main part of the paper: presentation and criticism of the alternative, agent-based account proposed by Michael McKenna.

4 Michael McKenna’s agent-based approach

According to Michael McKenna and some other philosophers who discussed the reasons-responsiveness account, e.g., Carl Ginet, what should be reasons-responsive is the agent herself, and not any mysterious mechanism. For McKenna, the first condition of moderate reasons-responsivity – the receptivity condition – is defined in the same way as by Fischer and Ravizza for the mechanisms, but the second condition – the reactivity condition – is modified so that it can be allowed to be reactive without being able to react otherwise.

How can we make sense of the idea of reactivity without being able to react otherwise? The proposal owes its plausibility to the recognition that, in the Frankfurt examples, there is an important difference between scenarios in which an agent recognizes sufficient reasons to act otherwise and *would act on them if there were no intervener* and those in which his recognition of reasons does not even motivate him to (try to) act on them. The difference is that in some scenarios, the intervener must intervene; and because of this, in those scenarios, the agent is not the source of her actions; she does not cause her action in a proper way. McKenna justifies his interpretation of theory by referring to the type of conditional used by new-dispositionalists:

(RR1): If S were to become aware of a reason R (to act otherwise than X) and if S retained during the relevant duration of time intrinsic agential properties P1-Pn, and if S were not interfered with in a way that would impede the casual efficacy of those properties, then S would not do X.

Although McKenna admits that this conditional may not account for the possibility to do otherwise, he claims that it still says something important about the agent herself: specifically, it says that the agent, just as she is, is reasons-reactive even if she cannot react differently.

Now I will present and advance the problem for agent-based theory that is mentioned but not developed by McKenna in his 2011's "Reasons-Responsiveness, Agents and Mechanisms". (McKenna, 2013: 170–176)

5 The problem of intrinsic undermining-possibility factors

Let there be – instead of some counterfactual external intervener – some property of the agent herself that does not play a part in the actual scenario yet that would prevent her from acting otherwise if she tried or wanted to do so. Specifically, consider the following scenario:

Anna has just become a student of philosophy and she is deliberating whether she should go to the first exam on her faculty. She is well-prepared and she should pass it without problems; more than that: she knows that if she does not go, she will fail the exam and thus lose her scholarship, which would in turn cause her mother, who is already taking financial care of five of her other children, a great distress. But she is just not in the mood and she decides to watch the movie instead of going to the exam. However, Anna suffers from a serious phobia, which she (nor anyone else) has discovered yet: she is extremely terrified of taking any exams at the University. In any scenario in which she would be just about to make a decision to go, she would become extremely frightened and, in the end, decide to stay home anyway.

It seems that, in the actual scenario, Anna is responsible not only for her decision not to go to the exam, but also for not going there, and therefore for causing her mother's distress. It can be shown, that McKenna's and Fisher's versions of reasons-responsiveness theory yield different results here, and that the result given by mechanism-based theory seems to be the proper one, in contrast to the result which we obtain if we choose the agent-based theory instead. McKenna's account fails in two ways. First, the justification in terms of new-dispositionalists conditional is no longer available, since the conditional is just false:

(a) If we assume that causing her mother distress is a sufficient reason for Anna not to make the decision she makes, it fails just because of the facts obtaining in the actual world, those facts being:

(a.1) Anna's recognizing a sufficient reason for not making the decision not to go to the exam and Anna's retaining all of the relevant agential properties during the relevant period and

(a.2.) Anna's making the decision not to go to the exam.

Since (a.1) makes the antecedent of the material conditional obtaining in the actual world true, and (a.2) makes the consequent of this conditional false, the whole material conditional is true. And since according to the most probable account of counterfactual conditionals, provided by D. Lewis, if a material conditional is false in world *w*, the associated counterfactual is also false in *w*, we can conclude that the new dispositionalist conditional is false.

(b) If we assume that Anna is just not such a decent person and causing her mother's great distress does not count as a sufficient reason for her to refrain from making the decision that causes the distress, we have to consider different possible worlds. Assume that Anna would recognize it as a sufficient reason if her mom were to have a heart attack as a result of her action. We then obtain a following conditional:

If (b.1) Anna was to become aware that her mom would have a heart attack if she decides not to go to the exam, and (b.2) she retained relevant agential properties P1-P_n during that time and (b.3) if she was not interfered in a way that would impede the causal efficacy of those properties, then (C) Anna would decide to go to the exam.

This conditional is, again, false. If the antecedent (b.2) were true, Anna would also retain the property of having a terrible phobia of taking exams at the University. This property would activate when Anna would be just about to decide to go to the exam and prevent her from making that decision, therefore causing (C) to be false. The truth of (b.3) is not impeded at all, it is just because of the causal efficacy of some of the properties she has that she fails to make the decision.

Also, Anna is not reasons-responsive in a sense specified by McKenna, because she fails to be reasons-reactive even on the weakened reading of reactivity. Let R-R_y be the range of reasons which Anna would recognize as sufficient, and R1-R_n the range of reasons that would motivate Anna to (try to) do otherwise. Anna would react to those *if she hadn't suffered from* phobia; in contrast, she wouldn't react to the reasons from the range R_n-R_y regardless of whether she suffers from phobia or not.

Unfortunately, this time the difference is not enough, at least if we want to claim, as McKenna did, that the key feature of the difference is that, in some of the scenarios, the agent is the ultimate cause of her actions. Both in the scenario in which the reasons from the range R1-R_n are present and in

those where Anna is confronted with reasons from Rn-Ry range, it is her – and only her – who or what is the cause of her actions.

On the other hand, according to Fischer and Ravizza's mechanism-based theory, Anna is responsible for both her decision and her mother's distress. Since responsibility for the action itself follows on their account from the responsibility of the action's consequences, I will only show how this account grants Anna's responsibility for her mother's distress. Keeping in mind the pertinent problem of mechanisms individuation, let me vaguely specify the mechanism playing this role in Anna's making the decision not to go to the exam and following this decision as "more or less rational deliberation concerning going to the exam". Now, the first condition for Anna's responsibility is fulfilled: there certainly are possible scenarios in which this (type of) mechanism operates and leads Anna to different action than staying at home watching TV. Also, the second condition is satisfied. There are other bodily movements, constituting Anna making the decision to go to the exam, which, other things being equal, would lead to a situation in which Anna's mother would not be distressed.

Fischer's result – that Anna is responsible for her decision, her following the decision, and her mother's distress – is certainly the more intuitive one. However, without going into any details, McKenna points at two possible ways of defending his theory. I will now consider his proposals and give some arguments against treating them as good ways of proceeding with establishing the plausible account on moral responsibility.

6 (Im)possible solutions to the problem of intrinsic undermining-possibility factors

1.

One option that McKenna finds at least worth considering is to walk away a bit further from the common intuitions and simply accept the conclusion that his account yields in problematic cases. But this proposal should not be pursued. Those are not only the reasons which have to do with the incentive to stay as close to intuitions as possible that witness against it; there are normative reasons as well, and they seem to be even more important.

Cases suggested by McKenna are much more down-to-earth than those usually employed in Frankfurt-style examples, and they find application in stories that can be easily encountered in real life. The agents in question suffer psychological impairments that do not allow them to perform some particular actions; although they can freely perform some actions which fully constitute not-performing that actions. If we accept McKenna's result in case of Anna, we will have to accept that

in every situation of this type the mentally impaired agent is not responsible for at least those consequences of her actions that follow from the omission that her actual-scenario action constitutes. Also, we should accept that if an agent performs some positive act – where she performs it on the basis of her factual desires, beliefs and so on – such that she would perform that act anyway, in any differing circumstances, she is not responsible for the act. If we do not simply adopt incompatibilist intuitions, according to which an agent is never responsible for the act which she could not have avoided in the first place, there seems to be no convincing reason why we should adopt this conclusion in the case of agents suffering from some hidden phobias or diseases. To the contrary, it seems that we just should not do that. Being a proper agent of moral assessment is usually strongly connected with personhood, and if we want to claim that in the types of situations discussed – among which we should count inactive phobias, psychosis in regress and some other mental issues that are temporarily not active – we are moving dangerously close to excluding people suffering from mental illnesses from the sphere of moral agents, and thus, plausibly, from subjects that we would call “persons”.

However, it is worth noticing that if we wanted to change our minds and accept that we should not hold Anna ultimately responsible, then it would constitute an independent reason against Fischer’s version of the theory.

2.

The other undeveloped solution which McKenna points at, is to somehow exclude properties constituting psychoses, phobias, and other mental impairments of the relevant kind from those which constitute the agent herself. Although, in contrast with the first proposed solution, this one seems more reasonable from a wider perspective, which deals with the issue of who should be regarded as a moral agent, it does face a technical difficulty: it simply looks like a step back towards some variation on mechanism-based theory. The key problem is that we cannot simply specify some range of conditions, say ‘suffering from phobia or psychosis’, which we would always exclude from the relevant agential properties when assessing agent’s responsibility for the performed action.

We cannot do that because in a lot of scenarios it would yield a seriously undesired result: that the person is reasons-responsive and thus, at least in most cases, responsible when the psychotic episode *is* present.

Consider Jane, who steals a bike belonging to a tourist who carelessly left it by the store without any protection. Jane took the bike because she was strongly convinced that it is a gift for her from

Saint Mother Mary: Jane is a very religious person with schizophrenic tendencies; she was praying relentlessly over the last few weeks for a new bike, and when she entered into the psychotic episode she just came to believe that the first bike she saw is the fulfilment of her prayers.

But now, if we insist on the exclusion mentioned earlier, Jane seems to be perfectly reasons-responsive in performing her action. There are certainly some reasons that Jane would both recognize and act on them in alternative scenarios in which she retains all her relevant properties; her psychosis, which played the key role in the actual scenario, would not be counted among them. Even if we do not want to treat reasons-responsivity as a sufficient condition of responsibility, and thus we may find a way to explain not holding Jane responsible for her action, it simply does not seem right – having in mind all that had been said so far about reasons-responsiveness – to say that she was reasons-responsive while stealing a bike.

On the other hand, if we try to specify the conditions in order to include some crucial features of the actual scenario, for instance: ‘suffering for phobia, which did not work while the agent was performing the action’, how far do we really move away from the mechanism-based account?

Faced with the abovementioned difficulties, I conclude that internal non-active undermining-responsibility factors are a serious threat for agent-based theories even if they can be convincingly defended against Frankfurt-type examples with external intervener.

7 Diana, the addict – the problem for both approaches?

Finally, I would like to discuss the example that may at the first sight seem to pose a problem for both mechanisms-based and agent-based versions of reasons-responsiveness theory.

Consider the scenario described below:

Diana, the recovered drug addict from a quite pathological family faces serious pressure to take a drug during Christmas family meeting. However, although it requires a lot of effort and some suffering on her side, she keeps refusing it and her abstinence is left intact. But Diana is really devoted to her new, sober lifestyle. In fact, she is so devoted that neither the mechanism leading to her refusal nor she herself as an agent is reasons-responsive when it comes to this particular topic. Anyway, it seems proper to think that Diana is responsible – in this case, praiseworthy – for keeping her abstinence.

In order to be able to take intuitions concerning Diana’s responsibility seriously, we have to say in more “down-to-earth” terms what it would really mean for our imagined picture of Diana’s act if

either the act was not the effect of a reasons-responsive mechanism or if Diana herself were not a reasons-responsive agent.

According to Fischer-Ravizza theory, there are few ways in which the mechanism can fail to be reasons-responsive. I will now consider all these options in the context of the discussed example. First, and again despite of the prominent problem of mechanism-individuation, the mechanism in question has to be at least vaguely specified. What I propose here to consider, very much in the spirit of Fischer's own candidates for relevant mechanisms, is the mechanism of rational deliberation. Some details will be added in the following discussion of the available options.

(1) There are **no** different possible worlds in which the same kind of mechanism operates in S and S has a different sufficient reason to do otherwise and S recognizes this reason.

This option is not really appealing in its most simple formulation; however, we can try to defend it by proposing quite complicated scenarios. In its most probable version, it emerges if we assume either that staying sober is the highest point in Diana's priority list or, even better, that *never to decide to take a drug* is such a goal. In the first case, in order for us to have option (1) as undermining reasons-responsiveness, we need Diana to be insensitive to any reasons, including such reasons as 'if you decide not to take a drug at t1, you will take the same dose of a drug at t2, t3 and again at t4; while if you decide to take a drug at t1, it will be for the last time and you will then remain sober for the rest of your life'. However odd the scenario in which Diana is faced with such a complex reason is, it has to be taken into account, and if Diana is supposed not to recognize this reason and any other reason of this structure as sufficient to take a drug at t1, then her lack of recognition stays in direct contradiction with what is her ultimate goal. And in such a situation it seems very implausible to maintain that the mechanism leading to her decision is any kind of "rational" deliberation. And even if we are eager to weaken the specification of the mechanism to "just deliberation" (whatever that is supposed to mean), the kind of deliberation that she performs here seems to be far too irrational to be able to grant moral responsibility; in fact, it seems to simply undermine it. The possible way out here is to suppose Diana's most important goal is to never violate the following rule: *never decide to take a drug on your own*. Now it is disputable whether we can postulate the scenarios in which Diana is faced with a reason of the form: 'if you don't (on your own) decide to take a drug at t1, you will make (on your own) more such decisions at t1, t2, t3 etc. In what sense are these latter decisions Diana's own if she is faced with such a reason in advance? To defend this way of lack of Diana's reasons-responsiveness in the discussed scenario, we must be ready to answer this question.

(2) There are some different possible worlds in which the same kind of mechanism operates in Diana and she has a different sufficient reason to do otherwise, but there is no understandable pattern of her recognition of those reasons.

This is the option that I do not find appealing at all, and I mention it just to provide the full spectrum of options that can be considered. If the pattern formed by Diana's recognitions is not intelligible at all (especially if we remember that it has to be intelligible on the assumption of the agent's values and reasons and not from the point of view of the observer who does not have access to those), then the agent acts in a way which is too irrational to consider her sane in any plausible sense; in fact she seems to act just randomly. And yet, in the present discussion, I want to grant the agent at least some degree of rationality in her acting: I want the mechanism not to be reasons-responsive, but I don't want the agent to be a total lunatic. Such a picture of her is provided by the two options that follow.

(3) There are some different possible worlds in which the same kind of mechanism operates in S and S has a different sufficient reason to do otherwise and S recognizes this reason, and her recognitions form an intelligible pattern, but there are no moral reasons among those which S would recognize.

This option can be easily construed from already considered options. Assume that Diana recognizes, in a coherent manner, all the reasons of the type discussed in option (1). However, she is completely blind to moral reasons that we would think should make her violate her strong resolution. We have to therefore accept that Diana would not recognize, for example, the reason: '(only) if you don't take the drug, the nuclear bomb will be dropped on some village and all those innocent villagers will be killed.' as sufficient to take a drug. Or to be even more direct: '(only) if you don't take the drug, almost all of people in the Earth will die in a terrible suffering within 5 hours from you making your decision.' It seems that these are the reasons that should be taken into account when one makes some decision concerning her own actions. However, it is not obvious at all that inability to recognize such reasons makes one not morally responsible for acting in particular circumstances when no such reasons are present.

(4) Diana recognizes sufficient reasons, including some moral ones, in different possible worlds in which the same kind of mechanism operates in her, but there is no possible world in which the same mechanism operates and Diana acts for such a reason. In other words, Diana is moderately reasons-receptive, but she is not even weakly reasons-reactive.

Such a possibility was analysed and rejected by Fischer (Fischer and Ravizza, 1998: 89–90) during the discussion of quite a similar type of example. Fischer asks what would happen if the individual had made a decision to never act on any sufficient reason concerning a particular action, and he answers to himself that we just should not keep that decision fixed, i.e., we should not regard it as a part of mechanism that led to action. The question here is whether we *can* – coherently and without begging the question – claim that, in our scenario, it is at least possible to keep the decision fixed and still hold Diana morally responsible. Also, it should be done in such a manner that the only clear principle postulated by Fischer and Ravizza for mechanisms – the principle that the description of the mechanism does not entail a particular action by itself – is not violated.

If we are ready to accept any of those options as possibly true of the agent who we still think is morally responsible for her action, we are provided with a counterexample to Fischer and Ravizza's mechanism-based theory. What about the agent-based account? Recall the conditional supporting it. Filled with the details of the discussed scenario, it says:

If Diana were to become aware of a reason R (to take the drug) and if Diana retained during the relevant duration of time intrinsic agential properties P_1 – P_n , and if Diana were not interfered with in a way that would impede the casual efficacy of those properties, then Diana would take a drug.

Does the falsity of this conditional imply that Diana is not responsible for not taking the drug in the actual scenario? I see no good reason for such inference.

Furthermore, according to McKenna, if Diana is not reasons-responsive with respect to action X, it is either true that she does not recognize any sufficient reasons for not performing action X, in which case we can apply considerations from the preceding paragraphs, or she is not reactive. Following McKenna's approach, not being reactive means that there are no scenarios in which Diana would react otherwise if there were no counterfactual interveners or any other intervention than can be perceived as external to her, preventing her from reacting. But nonexistence of such scenarios is exactly as probable as lack of reactivity in Fischer's sense. It seems that, in this type of example, agent-based and mechanisms-based theories stand or fall together.

8 Conclusion

Two variations on reasons-responsiveness theory: the mechanisms-based account and the agent-based account were compared on the grounds that abstract from the most commonly discussed problem of mechanism-individuation. The problems that emerge for agent-based account when we

analyse particular scenarios were pointed at and developed, in favour of mechanisms-based approach.

The possible counterexamples to both accounts were presented and it was clarified what follows from accepting such type of counterexamples. It is left for the reader to decide whether it is worth to pursue it further or if it should be abandoned when vague intuitions supporting it are analysed and specified.

Odzivnost na razloge: Še en pregled vidika teorij osnovanih na akterju in teorij osnovanih na mehanizmu

Fischer in Ravizza trdita, da alternativne možnosti ravnanja niso nujni pogoj za moralno odgovornost. Avtorja prepričajo Frankfurt primeri in sta tako pripravljena zagovarjati stališče zaradi nerazrešenih problemov kompatibilnosti determinizma in možnosti drugačnega ravnanja. Namesto tega predlagata, da za pripisovanje moralne odgovornosti ljudem ti potrebujejo »nadzorno usmerjanje« (angl. *guidance control*), ki je odvisno od mehanizmov, ki vodijo do tega, da je dejanje zmerno odzivno na razloge. Koncept zmerne odzivnosti na razloge (*MRR – moderate reasons-responsiveness*) je razložen na naslednji način: vrsta mehanizma M je MRR če in samo če (1) v več možnih svetovih, kjer je mehanizem dejaven in ima S zadostni razlog, da ravna drugače, S prepozna takšen razlog kot zadostni; (2) S-jeva dispozicija, da prepozna različne razloge tvori razumljiv vzorec; (3) v vsaj enem izmed teh svetov S ravna v skladu s prepoznanim zadostnim razlogom. Ko sta avtorja predstavila svoj vidik v knjigi *Responsibility and Control. A Theory of Moral Responsibility* (1998), se je pojavilo veliko razprav in kritik. Ena izmed glavnih poant kritikov je bila nejasnost koncepta mehanizma in pomanjkanje kriterijev identitete za različne vrste mehanizmov. To je vodilo nekatere filozofe, med drugimi Carl Ginet in Michael McKenna, da so prevzeli teorijo odzivnosti na razloge osnovano na akterju. Ta sprememba je naletela na druge probleme. Prvo je bilo izpostavljeno, da jo ovržejo primeri vrste Frankfurt, ker akter, s katerim je protidejstveni posrednik povezan, ni odziven na razloge, čeprav je moralno odgovoren. S tem se Ginet in McKenna ne strinjata. Vendar pa kasneje priznata, da obstajajo možni primeri, povezani z akterju internim »posrednikom«, ki predstavljajo grožnjo na akterju osnovani teoriji. Eden izmed nerazvitih primerov, ki predstavlja takšen problem, je situacija, kjer duševno bolan akter ravna, ko njegova bolezen ni aktivna, vendar bi se ta aktivirala in preprečila drugačno ravnanje v alternativnih scenarijih. V tem članku predstavim primere, osnovane na predlogih McKenna in trdim, da predstavljajo veliko večji problem za teorijo osnovano na agentih kot McKenna prizna. Prav tako predstavim možni protiprimer tako teoriji osnovani na agentu kot teoriji osnovani na mehanizmu, ki zadeva okrevano odvisnico, ki zavrača droge, vendar se zdi, da je moralno odgovorna (v tem primeru hvalevredna) za njeno upiranje, čeprav ni odzivna na razloge na pravi način.

Ključne besede: inkompatibilizem, polkompatibilizem, odzivnost na razloge, moralna odgovornost.

Works Cited

- Fischer, J. M., Ravizza, M. (1998). *Responsibility and Control. A Theory of Moral Responsibility*. Oxford: Cambridge University Press.
- Frankfurt, H. (1969). "Alternate possibilities and moral responsibility". *The Journal of Philosophy*, 66(23), pp. 829-839.
- Ginet, C. (2006). "Working on Fischer and Ravizza's Account on Moral Responsibility". *The Journal of Ethics*, 10, pp. 229–253.
- McKenna, M. (2013). "Reasons-Responsiveness, Agents, and Mechanisms". In Shoemaker, D., *Oxford Studies in Agency and Responsibility, Volume 1*. Oxford: Oxford University Press.
- van Inwagen, P. (1976). "The Incompatibility of Free Will and Determinism". *Philosophical Studies*, 27, pp. 185–199.