Reasons-Responsiveness and Moral Responsibility:

the Case of Autism

**§1. Introduction**

Theories of moral responsibility can be separated, roughly and with exceptions, into three broad categories depending on which features or abilities of agents the theory takes to be central for ascriptions of praise and blame (or the appropriateness thereof). Quality of will theories hold that the matter of an agent’s responsibility for a particular action is decided by determining whether or not the action expressed an objectionable quality of will on the agent’s part. Mesh theories claim that agents are responsible for acts which flow from the agent’s “real self,” or for actions that come from an appropriate mesh between certain key aspects of her agency. Finally, reasons-responsive theories claim that agents are responsible for their actions insofar as they possess the rational capacity to recognize reasons for acting and to guide their actions in accordance with those reasons. The focus of this paper will be the latter sort of theory, and I will be primarily interested in John Martin Fischer and Mark Ravizza’s version of it as presented in their seminal work, *Responsibility and Control*. Theirs is, by far, the most influential such theory, and it gives the most comprehensive statement of the reasons-responsive approach to moral responsibility on offer. However, I think that it is open to an important challenge from the empirical literature on autism spectrum disorder (ASD). More specifically, I think that evidence that individuals with ASD show signs of a pervasive deficit in counterfactual thinking opens the possibility of counterexamples that give reason to think that Fischer and Ravizza’s construal of reasons-responsiveness fails to meet the necessary and sufficient conditions for moral responsibility as they claim.

The argument of the paper will proceed as follows: in the following section I will present the key features of Fischer and Ravizza’s reasons-responsive theory of moral responsibility. I will then present empirical evidence suggesting that individuals with ASD demonstrate a unique deficit in a particular type of counterfactual reasoning. Finally, I will argue that this deficit makes it possible that autistic individuals may meet all of Fischer and Ravizza’s conditions for moral responsibility yet fail to be morally responsible, thereby showing that the theory fails to give an adequate account of the conditions under which an agent is morally responsible.

**§2. Fischer and Ravizza’s Reasons-Responsive View**

The goal of the reasons-responsive theorist, most generally, is to give an account of the sort of control that is necessary for an agent to possess in order to be the apt target of praise and blame for her actions, and this sort of control, according to these theorists, consists in a particular rational capacity – namely, the capacity to recognize and respond to reasons. The task of the reasons-responsive theorist, then, is to give an account of what it means to be properly responsive to reasons and how such responsiveness grounds moral responsibility.

Fischer and Ravizza attempt to characterize the relevant sort of responsiveness by carving out a middle ground between what they call “strong” and “weak” reasons-responsiveness.[[1]](#footnote-1) A mechanism which issues in action in the actual sequence[[2]](#footnote-2) is strongly reasons-responsive, they say, if, in any alternate sequence, it “were to operate and there were sufficient reason to do otherwise, the agent would *recognize* the sufficient reason to do otherwise and thus *choose* to do otherwise and *do* otherwise.” (Fischer and Ravizza 1998, 41) It follows from this that a mechanism can fail to be strongly reasons-responsive if the agent fails to recognize the reasons that there are for acting, fails to make a decision which aligns with those reasons, or fails to act on that decision. Fischer and Ravizza refer to the first sort of failure as a failure to be “receptive” to reasons and to the latter two sorts as failure to be “reactive” to reasons. So, strong reasons-responsiveness requires that a mechanism allow an agent to recognize reasons and to translate them into action whenever sufficient reasons exist. However, Fischer and Ravizza argue that strong reasons-responsiveness is too strong a requirement for moral responsibility, for we can think of cases in which an agent is morally responsible but is not strongly reasons-responsive. Consider, for example, a weak willed agent who has a desire (though not an overwhelming one) to shoplift, knows that shoplifting is wrong, and yet decides to shoplift despite this knowledge. Given that this agent recognizes a reason not to steal but fails to act on that reason, it appears that he is not strongly reasons-responsive. However, if this is the case, then he is also not morally responsible for his shoplifting, but surely this is the wrong verdict.

One way of resolving this problem is to construe reasons-responsiveness as requiring something much weaker. Fischer and Ravizza consider this solution and define weak reasons-responsiveness as follows: an agent is weakly reasons-responsive if, holding fixed the mechanism that operates in the actual sequence, there exists “some possible scenario (or possible world) in which there is a sufficient reason to do otherwise, the agent recognizes this reason, and the agent does otherwise.” (44) However, this won’t quite do either. For we can imagine an agent who meets these conditions yet is intuitively not a responsible agent. Suppose, for example, that our shoplifter’s desire to steal is not such that he could resist succumbing to it except in one possible scenario. If, say, at the moment he were about to pocket the stolen item it had been the case that an angsty teenager was reading *Ham on Rye* at the café next door, then the shoplifter would have seen this as a reason not to steal and would have refrained from stealing. In this case, the shoplifter meets the requirements of weak reasons-responsiveness, but there is a strong intuition that he is not, on this basis, morally responsible. Fischer and Ravizza call this the “problem of strange patterns.”

So, strong reasons-responsiveness is too strong, and weak reasons-responsiveness is too weak. In order to carve out a middle ground, Fischer and Ravizza propose that what is needed is a kind of moderate reasons-responsiveness, and they develop this notion by appealing to an asymmetry between reasons-receptivity and reasons-reactivity. More specifically, what is required for moral responsibility is a mechanism that is weakly reactive to reasons but which is receptive to reasons in a stronger sense. They suggest that by understanding these as asymmetrical we can make sense of both cases of weakness of will and cases of strange patterns since in the former case the agent seems to be strongly receptive to reasons, and in the latter case despite the agent being weakly reactive to reasons she is not receptive to reasons in the stronger sense. Thus, moderate reasons-responsiveness can give a plausible explanation of our intuitive judgments about each of these cases.

In an effort to further characterize what moderate reasons-responsiveness consists in, Fischer and Ravizza write,

In judging a mechanism’s receptivity, we are not only concerned to see that a person acting on that mechanism recognizes a sufficient reason in one instance; we also want to see that the person exhibits an appropriate *pattern* of reasons-recognition. In other words, we want to know if (when acting on the actual mechanism) he recognizes how reasons fit together, sees why one reason is stronger than another, and understands how the acceptance of one reason as sufficient implies that a stronger reason must also be sufficient. (1998, 70-71)

A mechanism that shows this sort of pattern, Fischer and Ravizza call “regularly receptive.” In order to determine whether or not a mechanism possesses such a pattern, they ask us to imagine a third party conducting an imaginary interview where the agent is asked various questions about actual and hypothetical scenarios. If, they say, the agent’s answers to questions in the imaginary interview produce a pattern of receptivity that is understandable by the third party, then the mechanism is regularly reasons-receptive. In short, on their view, “Regular reasons-receptivity … is reasons-receptivity that gives rise to a minimally comprehensible pattern, judged from some perspective that takes into account subjective features of the agent (i.e., the agent’s preferences, values, and beliefs)…” (73) Furthermore, they claim, this pattern must be one that is “minimally grounded in reality.”

Reactivity, on the other hand, requires no such pattern. Rather, if it is the case that a mechanism is reactive to reasons in *some* possible scenario, then that mechanism is reactive to reasons in *any* scenario. This is because, to use Fischer and Ravizza’s phrase, reactivity is all of a piece. Given this, they summarize the asymmetry between receptivity and reactivity as follows:

In the case of receptivity to reasons, the agent (holding fixed the relevant mechanism) must exhibit an understandable pattern of reasons-recognition, in order to render it plausible that his mechanism has the “cognitive power” to recognize the actual incentive to do otherwise. In the case of reactivity to reasons, the agent (when acting from the relevant mechanism” must simply display *some* reactivity, in order to render it plausible that his mechanism has the “executive power” to react to the actual incentive to do otherwise. (75)

So, on their view, an agent has control over her actions in the sense required for moral responsibility, only if those actions flow from a mechanism that is moderately reasons-responsive in the sense just described.[[3]](#footnote-3)

**§3. The Challenge from ASD**

This view of responsibility is appealing for a number of reasons and has been widely held as, perhaps, the most plausible account on offer in the literature. However, the view is open to an important challenge from the empirical literature on autism. As I noted in §1, empirical evidence suggests that individuals with ASD display important deficits in counterfactual thinking. These deficits are such that they have implications for both the cognitive ability to recognize certain kinds of reasons as well as for a number of executive functions. This feature of ASD has important implications for Fischer and Ravizza’s characterization of moderate reasons-responsiveness which, recall, relies on an important asymmetry between reasons-receptivity, a cognitive power to recognize reasons, and reasons-reactivity, an executive power to react in light of reasons. The empirical data regarding ASD suggests, I will argue, that the receptivity/reactivity profile of individuals with the disorder is quite different than that of neurotypical individuals, and it is not clear that the account of that profile given by Fischer and Ravizza can accommodate individuals with ASD. In this section, I will briefly survey the empirical literature on counterfactual thinking on ASD and will then attempt to show how this literature poses a problem for the moderate reasons-responsive view that Fischer and Ravizza offer.

*§3.1. Counterfactual Thinking in ASD*

Autism Spectrum Disorder is a complex, developmental disorder that is, despite the vast amount of research dedicated to it, not well understood. As a result, over the past few decades, several cognitive theories of ASD have been proposed in an effort to explain the various features which characterize it. For example, it is claimed by many that individuals with ASD have pervasive deficits in their ability to impute mental states to themselves and others (often referred to as having a theory of mind) and that this deficit explains the behavioral features of ASD.[[4]](#footnote-4) Others have claimed that the primary deficit in ASD is a deficit in executive function ability.[[5]](#footnote-5) Still others have suggested that the core deficit in ASD is not a deficit at all, but a type of processing bias in which individuals with ASD process information in a piecemeal fashion, focusing on details at the expense of forming a coherent picture of how they fit together (this view is known as the weak central coherence thesis).[[6]](#footnote-6) Each of these hypotheses aims to give a unifying picture of ASD, but each of them falls short in various respects.[[7]](#footnote-7)

One promising way of understanding these various theories of ASD, it seems to me, is to think of them in relation to the capacity for counterfactual thinking as each of the processes that these theories focus on seems to include important counterfactual elements.[[8]](#footnote-8) Counterfactual thinking, fundamentally, involves suspending one’s presently observed reality and thinking about states of affairs that do not match that reality. (Harris, German, and Mills 1996) Humans, of course, use this sort of thinking quite often in everyday life for a range of reasons, including reasoning about possible causes of events, possible outcomes of actions, or even for fantasizing about some imagined reality. Importantly, this type of reasoning involves two separate components. The first is a representational component in which the individual represents some counterfactual state of affairs, and the second is an inferential component in which the individual draws inferences from the represented counterfactual state. The evidence on this type of thinking in ASD suggests that the latter component is intact but that there is a significant impairment in the former, representational ability.

To this end, Leevers and Harris (2000) showed that individuals with ASD do not demonstrate deficits in deductive, syllogistic reasoning in which the syllogisms contain counterfactual premises.[[9]](#footnote-9) Additionally, Peterson and Bowler (2000) found that autistic subjects were as adept as control groups at performing subtractive counterfactual reasoning by observing subjects’ ability to correctly answer questions of the form, “If X had not happened, then what state of affairs, Y, would have occurred?” So, based on these two studies, it seems as though children with ASD are able to successfully make inferences from counterfactual propositions or states of affairs.

Interestingly, in the Peterson and Bowler study participants were also asked to complete a standard false belief task,[[10]](#footnote-10) and of those participants with ASD who passed the counterfactual question only 44% were able to pass the false belief test as well. (compared to 76% in the typically developed control group). Peterson and Bowler explained these results by hypothesizing that the false belief task requires the spontaneous generation of the counterfactual state of affairs while the counterfactual state of affairs is stated explicitly in the counterfactual reasoning tasks. If this is true, then it suggests that there may be a deficit in the representational component of counterfactual thinking. In a more recent study, Grant, Riggs, and Boucher (2004) replicated the Peterson and Bowler results but added to that study by including non-standard false belief tasks. In these tasks, the false belief of the character in the story was made explicit[[11]](#footnote-11) so as to test the Peterson and Bowler hypothesis that the explicit features of the counterfactual reasoning tasks were the reason for the improved performance of individuals with ASD as compared to the standard false belief tasks. Altering the experiment in this way proved successful, and Grant et al. summarized their results as follows:

[I]t appears that autistic children’s difficulties in passing standard false-belief tasks derive from the cognitive requirements of the tasks and not from a flawed conceptual understanding of belief. Moreover, the cognitive difficulties that [children with autism] experience on standard false-belief tasks are associated both with the particular cognitive demands of counterfactual conditional reasoning and with drawing inferences/generating propositions where critical information is not made explicit. (184)

What all of this suggests is that individuals with ASD are able to reason about counterfactual states of affairs when such reasoning is either deductive as in the Leevers and Harris study or involves states of affairs in which the counterfactual components are made explicit. When the counterfactual elements are not made explicit, however, individuals with ASD show clear deficits in counterfactual thinking.[[12]](#footnote-12) The most plausible interpretation of this data, I think, is that it shows that individuals with ASD are impaired in the ability to represent counterfactual states when not provided with their component parts, an explanation supported by the findings of Grant, Riggs, and Boucher. If this interpretation is correct, then it could have far reaching implications for how we understand the cognitive basis for a number of autistic traits.[[13]](#footnote-13) However, for the purposes of this paper, the central insight to be drawn from this portion of the ASD literature is simply that there is, indeed, a deficit in counterfactual thinking in ASD, and that this deficit is, plausibly, one of the central features of the disorder. The question, to which I will now turn, is what are the implications of this deficit for a reasons-responsive theory like Fischer and Ravizza's?

*§3.2. Receptivity and Reactivity in ASD[[14]](#footnote-14)*

Fischer and Ravizza claim that a mechanism’s being moderately reasons-responsive (and belonging to the agent in the relevant sense) is both necessary and sufficient for the sort of control required for moral responsibility. The challenge from ASD that I will present here denies the sufficiency of this condition by claiming to show that there may be individuals with ASD who are moderately reasons-responsive yet fail to be responsible for a range of actions. The reason that ASD presents a unique challenge to this notion, I contend, lies in the cognitive/executive construal of responsiveness that Fischer and Ravizza propose and the unique deficits that individuals with ASD experience in both cognitive ability and executive control.

To see how this may be, consider the moral deliberation mechanism of a high-functioning autistic agent. The first question that must be asked is whether or not the mechanism is regularly receptive to reasons in the sense that Fischer and Ravizza’s view requires. That is, does the mechanism give rise to an understandable pattern of receptivity that is minimally grounded in reality? There is no evidence to suggest that individuals with ASD fail to meet this condition, and there is plenty of evidence to suggest that individuals with ASD show no impairments in their ability to recognize moral reasons. For example, individuals with ASD tend to be able to make the moral/conventional distinction at least as well as neurotypical agents.[[15]](#footnote-15) In other words, they seem clearly able to tell the difference between violations of moral rules and violations of convention, and this suggests that individuals with ASD are sensitive to moral considerations in an important way. Crucially, individuals with ASD are also able to draw the moral/conventional distinction in cases involving authoritative permission. That is, they are able to make judgments of the following type: “Moral harm, *x*, would be wrong even if authority figure, *A*, said it was permissible.” (Zalla et al. 2013) This ability is especially significant given that some have suggested that the presence of second-personal authority is the key distinguishing feature in the moral/conventional task. (Shoemaker 2011)

While the moral/conventional task has its problems,[[16]](#footnote-16) the fact that individuals with ASD can make the moral/conventional distinction suggests, at least, that autistic persons are no worse off in this respect than are neurotypical agents. So, the fact that individuals with ASD perform at control levels on this task suggest, I think, that they are receptive to moral considerations in an important way.

That autistic persons can make moral judgments in this more minimal sense, however, is not to say that the ability to make certain fine-grained moral judgments is unimpaired. Despite being able to recognize the moral/conventional distinction, individuals with ASD tend to have problems with other features of moral judgment. For example, studies have found that these individuals have trouble distinguishing intentional from unintentional harms[[17]](#footnote-17) as well as having problems distinguishing attempted harms from neutral actions. In other words, studies suggest that individuals with autism have a limited ability to incorporate information about the agent’s mental state in their judgments of a given action. This is, perhaps, not surprising since inferring another person’s mental state from his or her actions requires a robust counterfactual representation ability which, as we have seen, is impaired in ASD.

Importantly, many studies on moral judgment in ASD show that individuals with ASD have difficulties in offering justifications for their judgments. (Loveland et al. 2001; Zalla et al. 2013) In many cases, the explanation for any given moral judgment relies on an appeal to a general moral rule (e.g. “It’s bad to lie.”). This is, of course, consistent with the suggestion above that individuals with ASD rely more heavily on model-free processes of moral judgment, and it pairs well with general observation that individuals with autism often develop complicated compensatory heuristics that help to make up for some of the cognitive or emotional impairments associated with autism. In order to make sense of their surroundings and to function in the world around them, individuals with ASD must come up with a way to relate to the world in spite of the psychological factors associated with ASD. One way that they are able to do this is by identifying a set of rules by which they can orient their interactions with other people. As a result, individuals with ASD are often seen as rigid rule-followers,[[18]](#footnote-18) and this would explain the self-reported justifications for moral judgments like those just mentioned.

One might be concerned that these atypical features of moral judgment in ASD are such that they preclude autistic agents from meeting Fischer and Ravizza’s criterion for regular receptivity. After all, it matters to us whether or not an agent had a particular set of intentions in acting or whether she displayed good or ill will, and insofar as failing to be receptive to these features may amount to failing to be receptive to a particular set of moral reasons, it could be that individuals with ASD fail to meet the regular receptivity requirement. However, this is not the case. The fact that individuals with ASD are able to develop compensatory heuristics for moral judgments and to distinguish moral from conventional norms most of the time is evidence that they are capable of recognizing an understandable pattern of reasons that are minimally grounded in reality, and this general capacity is all that is required for regular receptivity according to Fischer and Ravizza. What this brings out, it seems, is an ambiguity in what Fischer and Ravizza take to constitute a moral reason. Fischer and Ravizza are explicit that their aim is to give an account of what it means to be responsible for a given *action*, and the fact that individuals with ASD make moral judgments in an atypical way does not preclude them from being regularly receptive to reasons so long as the judgments that are made yield coextensive sets of actions. Fischer and Ravizza's claim is *not* that regular receptivity requires making neurotypical moral judgments. Rather, it is that individuals must be receptive to reasons in such a way that would lead others to judge that it would be appropriate to respond to the individual with one or more of the reactive attitudes on the basis of the actions that flow from the reasons she recognizes. It may well be the case that what individuals with ASD are doing in these atypical cases is not moral judgment, but something just as good. Nonetheless, insofar as this form of judgment yields actions that are consistent with those that flow from neurotypical moral judgments, there is no basis on Fischer and Ravizza's view for denying that individuals with ASD fail to meet the regular receptivity requirement.[[19]](#footnote-19)

Despite all of this, given the cognitive impairments present in ASD, surely there are cases in which individuals with autism are regularly receptive and weakly reactive to reasons yet are not responsible. This is because all that is required for regular receptivity, according to Fischer and Ravizza, is that a third party giving an imaginary interview could identify a regular pattern of receptiveness to reasons that is minimally grounded in reality. Indeed, part of what distinguishes moderate reasons-responsiveness from strong-reasons responsiveness is that, on the moderate view, the agent is responsible for actions that actually flow from a mechanism with a general capacity for regular receptivity and weak reactivity to reasons rather than being responsive to the actual reason that there is for acting.[[20]](#footnote-20) It is uncontroversial that such a pattern exists in high-functioning autistic individuals. Indeed, the compensatory systems developed by individuals with ASD seem to serve precisely this function. However the evidence of impairments in counterfactual thinking described above suggests that there is a range of reasons to which individuals with ASD may not be receptive, namely those reasons that require a robust ability to represent counterfactual states. It follows from this, then, that it is possible that individuals with ASD may act from a mechanism that is regularly receptive to reasons yet is not receptive to the reasons that there are to perform, or not perform, a given action if those reasons require an ability to think counterfactually.[[21]](#footnote-21) An example may be helpful here.

Suppose that Adam is a high-functioning autistic agent who possesses, to a substantial degree, the counterfactual deficits described above. He recognizes norms of politeness and routinely follows them. Additionally, he recognizes that he has moral reason not to do or say things that cause others pain, and he regular acts in accord with such reasons. While having lunch at a local restaurant, Adam encounters a friend of his. Recognizing that it is polite to ask one’s friends about their family, he asks his friend if her brother is well. Let us suppose, however, that the friend’s brother is currently serving an extended prison sentence (a fact which Adam knows) and that this is a source of deep shame for Adam’s friend and her family. Adam’s inquiry is very upsetting to his friend, and her feelings are hurt. It seems to me that Adam is not morally responsible for hurting his friend in this case, and the reason for this is that, given his cognitive deficits, he would be unable to represent the counterfactual, “If I were in my friend’s situation, I would be ashamed of my brother, and it would be hurtful for others to bring this up.” This counterfactual representation, however, is what supplies the moral reason to refrain from acting in this case, and, so, lacking access to it seems to preclude Adam from being held blameworthy. Nevertheless, the mechanism which results in his action seems to meet Fischer and Ravizza’s criteria for regular receptivity, given its normal functioning in other situations, even though it fails to respond to the appropriate reasons in this case.[[22]](#footnote-22)

What is interesting about the case of ASD, however, is that the cognitive deficit that undermines receptivity for those with ASD undermines reactivity as well. This is trivially true in many cases (clearly, one cannot react to a reason that one does not recognize), but it is true in a wider range of cases than these as well. The reason for this lies in the nature of moral judgments more generally. Over the last decade, research in moral psychology has revealed that moral judgment appears to take place according to a dual process model. This dual process model has typically construed our moral judgments as occurring as a result of either an emotional or cognitive process.[[23]](#footnote-23) On this view, cognitive processes tend to result in consequentialist moral judgments whereas emotional processes tend to result in deontological moral judgments. A dual system view such as this is helpful in explaining certain peculiarities in our moral judgments. So, for example, in Judith Jarvis Thomson’s famous Bystander and Footbridge variants of the trolley problem, (Thomson 1985) individuals tend, overwhelmingly, to make utilitarian and deontological judgments, respectively. If a dual-process model of moral judgment is correct, this difference in judgments can be easily explained since, it is posited, the latter case activates an emotional system that the former does not.

In more recent studies, however, a different sort of dual process view has been proposed which makes use of a model-based vs. model-free framework.[[24]](#footnote-24) The central claim of this sort of view is that moral decision making can be best understood in computational terms. So, in some circumstances, moral decisions are made according to a representation of a model, or decision tree. The agent represents a model and then follows that model in order to reach some end goal. Decisions are then made according to whether or not certain courses of action conform to our model or help to achieve the desired end. In other cases, however, assigning model-free process is utilized in which positive or negative values are assigned directly to particular actions by way of positive and negative feedback which has been habituated over time. Actions which receive consistent positive feedback eventually take on the form of a moral rule. Fiery Cushman summarizes the distinction as follows:

Goal-directed actions require a working model of the world. You pick a desirable outcome, and then form a plan to bring it about. Thus, they correspond to the class of model-based reinforcement learning algorithms. In contrast, habits are reactive stimulus-response pairings that are strengthened when followed by reward. Executing a habit does not require planning toward a valued outcome, and thus corresponds to the alternative class of model-free algorithms. (2015, 49)

On this view, model-based processes are said to issue in consequentialist moral judgments while the model-free system issues in deontological moral judgments since achieving desirable consequences requires having a working model of the world while adhering to moral rules does not.

The model-based/model-free framework, it seems to me, offers a plausible way of thinking about autistic moral judgment. As I argued in §3.1, ASD is characterized by an impairment in the ability to represent counterfactual states and that this impairment is central to the observed deficits in executive function in individuals with the disorder given that it is a necessary component of the ability to represent problems and make plans. If this is true, then we should expect individuals with ASD to make decisions predominantly by way of a model-free process since the model-based process requires an ability to construct a model of the world as well as a plan for achieving goals, and, indeed, this seems to be correct. The compensatory systems developed by individuals with ASD seem to be sophisticated model-free processes that these individuals are able to develop over time in order to navigate the world around them. That is, they seem to consist in a highly complex system of paired inputs and outputs. Through habituation, high-functioning individuals with ASD are able to identify certain types of appropriate responses to various situations, and, so, to develop a habituated, model-free algorithm that will aid them in responding to new situations which share features of previously experienced situations. This way of understanding deliberation and judgment in ASD meshes well with the observed repetitive and restricted behaviors and interests in ASD insofar as the model-free process favors rule-based behavior.

The reliance on model-free processes has implications for reasons reactivity in ASD as well. Since counterfactual thinking is *impaired* in ASD rather than *lacking* all together it is possible that there may be cases in which a model-based system is required for making a successful moral judgment, and the autistic agent recognizes this and is able to form a model. However, it may be that, due to the impairment in counterfactual thinking, the model is ill-formed and not such that it could render the right sort of decision. Alternatively, it may be that the autistic agent is able to construct a model that is reasonably well formed but is unable to formulate a plan for action according to that model and is, therefore, unable to react to the reasons that she recognizes. In either case, the autistic agent seems to offer a genuine counterexample to Fischer and Ravizza’s view of the sort of reasons reactivity that is required for moral responsibility. Recall that their claim was simply that to be reactive to reasons, a mechanism must be able to react to reasons in at least one scenario since doing so would demonstrate an executive ability that is all of a piece. However, what these considerations about moral judgment in ASD seem to show is that such judgments implicate two *kinds* of executive power and that it is possible that a mechanism may fail to issue in morally responsible action if it possesses one but not the other. If this is true, then Fischer and Ravizza’s claim that moral responsibility requires only weak reactivity seems to be false.

Michael McKenna an Alfred Mele have each separately made a similar objection to this portion of Fischer and Ravizza’s view. In each case, the author asks us to imagine an agent who is reactive to one reason yet seems to be operating from a mechanism that is intuitively not responsible.[[25]](#footnote-25) Importantly, neither of them takes this to be a reason to reject Fischer and Ravizza’s view. Instead, they suggest that the picture of reactive merely needs to be revised. As Mele puts it, “An attractive strategy for avoiding the (apparent) problem that I have been developing is to beef up the reasons-reactivity condition in such a way that … agents with … severe psychological maladies of the pertinent kind do not count as reasons-reactive enough to be morally responsible for the relevant behavior.” (2006, 290) In response, Fischer accepts this revision to his and Ravizza’s account of reactivity, saying,

This posits a more refined notion of moderate reasons-responsiveness, with what might be called “spheres of responsiveness;” the “outer spheres” would not necessarily indicate sufficient responsiveness for moral responsibility. Of course, it may not be straightforward to characterize precisely the “borders” of the spheres; that is, it might not be easy to say exactly what degree of strength of the relevant sort of urge renders the agent in question immune to moral responsibility. (2006, 328)

Despite this alteration, he maintains that this revision is consistent with thinking of reactivity as “all of a piece” since one way of interpreting this notion of spheres of responsibility is to say that it simply means that “an agent who can react to any reason may have great difficulty in doing so in any particular context.” (328 fn17) So, Fischer claims, accepting the objection from McKenna and Mele preserves the basic features of the Fischer and Ravizza view.

However, a similar response to the objection that I have presented here is not open to Fischer and Ravizza for at least two reasons. First, the empirical evidence regarding the dual process system of moral judgment shows that individuals react to reasons in two ways which are distinguishable from one another. On the basis of this evidence, then, the objection from ASD explicitly denies the “all of a piece” claim made by Fischer and Ravizza. Second, Mele’s objection to Fischer and Ravizza is presented in terms of irresistible desires as is Fischer’s response to it. Because of this, Fischer’s response offers a picture of agents as possessing concentric spheres of responsibility where the agent, in the inner spheres, is not confronted with desires or urges which block her ability exercise guidance control but, in the outer spheres, she may be. Thus, reactivity is a matter of degree, and, he thinks, we need not worry about those outer spheres where the agent is overcome by her urges and where our intuitions are fuzzy anyway. However, the objection from ASD rests not on the presence of overwhelming urges or on *degrees* of reactivity. Instead, it rests on the presence of an impairment in a particular *kind* of reactivity. Because of this, it makes little sense to talk of spheres of responsiveness where reactivity is weakened as one moves toward the outer spheres. Rather, responsiveness requires both kinds of reactivity regardless of which sphere one finds oneself in. So, the objection from ASD shows, I contend, that reactivity is not all of a piece and that, as a result, an individual can fail to be morally responsible while meeting Fischer and Ravizza’s conditions for reactivity, *even if these conditions are strengthened* in the way that McKenna and Mele suggest.

One response to this objection that Fischer and Ravizza might offer is this: the fact that moral judgment proceeds according to two separate processes suggests that there are two separate mechanisms at work, and so, insofar as individuals with ASD react to reasons on a model-free process but do not do so on a model-based process this suggests that their model-free mechanism is weakly reactive while their model-based mechanism is not. However, this is to oversimplify the dual process model of moral judgment. This is because the two systems are not simply independent systems which issue in particular judgments on their own. Rather, as Cushman puts it,

The emerging picture is not … as simple as “two independent systems: model-based and model-free.” For one thing, each of the “systems” clearly comprises multiple dissociable components … In addition, decision making in the model-free system appears to involve independent and opponent “go” and “no-go” processes. More critical to the current discussion is the evidence that model-free and model-based systems interact closely. The successful operation of each system largely depends on interactions with the other. (2013, 280)

So, given that Fischer and Ravizza define a mechanism as simply the process by which an action is produced, there is no empirical basis for treating model-based and model-free processes as distinct mechanisms since both processes will be involved in the production of a given judgment or action. My claim here is that individuals with ASD rely much more heavily on a model-free system than do neurotypical individuals, not that they operate entirely on a model-free system. If this is true, then it would be implausible for Fischer and Ravizza to claim that a different mechanism is at work. One way around this difficulty might be for Fischer and Ravizza to claim that mechanisms are to be understood much more specifically than this. That is, they might claim that the mechanism issuing in action is to be identified with the actual psychological details or brain states of the agent. In their book, they are quite vague about how precisely to identify the mechanism in question and to hold it fixed, but there are good reasons for them to avoid identifying mechanisms with actual states of the brain.[[26]](#footnote-26)

So, if Fischer and Ravizza’s account of weak reasons reactivity is not satisfactory what type of reactivity would do the trick? It seems to me that it would need to be something much stronger than what they suggest and considerably stronger than the type suggested by McKenna and Mele. What is needed, if we are going to get things right in the case of ASD is something that might be referred to as “regular reactivity,” to riff on Fischer and Ravizza’s terminology. Regular reactivity would involve a regular pattern of reacting to reasons which issues from a well-balanced dual process system of deliberation, i.e. a system in which the model-free and model-based systems are functionally typical and act in concert with one another. This requirement would avoid the problem of judging individuals with ASD to be responsible for actions when they lacked the capacity to react to the relevant reasons. However, accepting this version of reasons reactivity will prove problematic for Fischer and Ravizza.

I have argued that ASD seems to pose a challenge to Fischer and Ravizza’s account of moderate reasons-responsiveness in two ways. First, it is clear that individuals with ASD demonstrate regular patterns of receptivity to reasons yet are not responsible for a range of actions given that their cognitive deficits rule out their having access to a certain range of reasons. So, Fischer and Ravizza’s notion of regular receptivity is too weak and must be made stronger if it is to render the correct verdict regarding individuals with ASD. One way of doing this would be to claim, as Patrick Todd and Neal Tognazzini (2008) do, that Fischer and Ravizza should adopt a stronger account of regular receptivity which says that an agent must be receptive to the actual reasons that there are for acting.[[27]](#footnote-27) Second, given the empirical evidence for a dual process system of moral judgment together with the executive control deficits displayed by individuals with ASD, it seems to be the case Fischer and Ravizza’s account of reasons reactivity is also too weak insofar as it may also count those with ASD as responsible in cases where they seem clearly not to be. One possible way around this, I have just suggested, is to adopt a view which says that regular reactivity to reasons is necessary for moral responsibility, thereby making this condition of reasons-responsiveness considerably stronger as well. However, it seems to me that Fischer and Ravizza cannot take on both of these suggested revisions to their account. To do so would be to say that in order to be responsible an agent must be able to recognize the actual reasons there are for acting and then must regularly react to those reasons. It is not clear, though, that there is a relevant difference between this formulation and one which claims that responsibility requires the ability to recognize reasons, choose to act on them, and then to so act. Recall, however, that this just is what it means to be strongly reasons-responsive, and Fischer and Ravizza explicitly say that this sort of responsiveness cannot adequately characterize the sort of control required for moral responsibility. It seems to me, therefore, that individuals with ASD pose a challenge to Fischer and Ravizza’s view, and it is not clear that their account has the resources to overcome it.

**§4. Conclusion**

To conclude, the empirical evidence regarding ASD shows that both components of Fischer and Ravizza’s account of moderate reasons-responsiveness fail to provide necessary and sufficient conditions for the type of control required for moral responsibility. Moreover, the most plausible revisions to the view that would accommodate this evidence are such that they would transform the view into one that Fischer and Ravizza explicitly deny. That is, accommodating the empirical evidence from autism requires pushing the reasons-responsiveness account away from its moderate position and into a stronger position which is untenable. So Fischer and Ravizza are left with something of a dilemma. Their stated aim is to give an account of control that is in reflective equilibrium in the sense that it gives equal weight to our intuitions about clear cases as to theoretical principles. However, the challenge from autism shows that this equilibirium has not been achieved. Either Fischer and Ravizza must choose between a moderate account that cannot handle clear intuitions about people with ASD or a strong account that cannot handle intuitions about clear cases of, say, weak willed behavior. Both positions, it seems to me are untenable, and, given that Fischer and Ravizza’s theory is the most plausible reasons-responsive account on offer, I contend that these challenges give us reason to look to something other than reasons-responsiveness to explain what it means to be a morally responsible agent.

REFERENCES

Baron-Cohen, Simon, Alan M. Leslie, and Uta Frith. "Does the autistic child have a 'theory of mind'?" *Cognition*. Vol. 21. (1985): 37-46.

Begeer, Sander, Mark Meerum Terwogt, Patty Lunenburg, and Hedy Stegge. “Brief Report: Additive and Subtractive Counterfactual Reasoning of Children with High-Functioning Autism Spectrum Disorders.” *Journal of Autism and Developmental Disorders*. Vol. 39. (2009): 1593-1597.

Blair, R. James R. "Brief Report: Morality in the Autistic Child." *Journal of Autism and Developmental Disorders.* Vol. 26, No. 5. (1996): 571-579.

Buon, Marine, Emmanuel Dupoux, Pierre Jacob, Pauline Chaste, Marion Leboyer, and Tiziana Zalla. “The Role of Causal and Intentional Judgments in Moral Reasoning in Individuals with High Functioning Autism.” *Journal of Autism and Developmental Disorders.* Vol. 43. (2013): 458-470.

Channon, Shelley, Sian Fitzpatrick, Helena Drury, Isabelle Taylor, and David Lagnado. “Punishment and Sympathy Judgments: Is the Quality of Mercy Strained in Asperger’s Syndrome?” *Journal of Autism and Developmental Disorders*. Vol. 40. (2010): 1219-1226.

Channon, Shelley, David Lagnado, Sian Fitzpatrick, Helena Drury, and Isabelle Taylor. “Judgments of Cause and Blame: Sensitivity to Intentionality in Asperger’s Syndrome.” *Journal of Autism and Developmental Disorders*. Vol. 41. (2011): 1534-1542.

Crockett, Molly. “Models of Morality.” *Trends in Cognitive Sciences.* Vol. 17, No. 8. (2013): 363-366.

Cushman, Fiery. “Action, Outcome, and Value: A Dual-System Framework for Morality.” *Personality and Social Psychology Review*. Vol. 17, No. 3. (2013): 273-292.

\_\_\_\_\_\_\_\_. “From Moral Concern to Moral Constraint.” *Current Opinion in Behavioral Sciences*. Vol. 3. (2015): 58-62.

Cushman, Fiery, Liane Young, and Joshua Greene. “Our Multi-System Moral Psychology: Towards a Consensus View.” In *Oxford Handbook of Moral Psychology.* 47-71. Edited by John Doris. New York: Oxford University Press, 2010.

Dolan, Ray and Peter Dayan. “Goals and Habits in the Brain.” *Neuron*. Vol. 80. (2013): 312-325.

Fischer, John Martin. “Reply: The Free Will Revolution.” *Philosophical Explorations*. Vol. 8, No. 2. (2005): 145-156.

\_\_\_\_\_\_\_\_. “Responsibility and Control.” *The Journal of Philosophy*. Vol. 79, No. 1. (1982): 24-40.

\_\_\_\_\_\_\_\_. “Responsibility and Self-Expression.” *The Journal of Ethics.* Vol. 3. (1999): 277-297.

\_\_\_\_\_\_\_\_. “Responsibility and the Actual Sequence.” In *Oxford Studies in Agency and Responsibility, Vol. 3*. 120-135. Edited by David Shoemaker. New York: Oxford University Press, 2015.

\_\_\_\_\_\_\_\_. 1987. “Responsiveness and Moral Responsibility.” In *My Way: Essays on Moral Responsibility*. 63-83. By John Martin Fischer. New York: Oxford University Press, 2007.

\_\_\_\_\_\_\_\_. “The Free Will Revolution (Continued).” *Journal of Ethics*. Vol. 10, No. 3. (2006): 315-345.

\_\_\_\_\_\_\_\_. “The Importance of Frankfurt-Style Argument.” *The Philosophical Quarterly*. Vol. 57, No. 228. (2007): 464-471.

\_\_\_\_\_\_\_\_. *The Metaphysics of Free Will: An Essay on Control*. Aristotelian Society Monograph Series, Vol. 14. Cambridge, MA: Blackwell Publishers, 1994.

Fischer, John Martin and Mark Ravizza. *Responsibility and Control: A Theory of Moral Resonsibility.* New York: Cambridge University Press, 1998.

\_\_\_\_\_\_\_\_. “Responsibility and Inevitability.” *Ethics*. Vol. 101, No. 2. (1991): 258-278.

\_\_\_\_\_\_\_\_. “Responsibility for Consequences.” In *Perspectives on Moral Responsibility*. 322-347. Edited by John Martin Fischer and Mark Ravizza. New York: Cornell University Press, 2004.

Frith, Uta. *Autism: Explaining the Enigma*. Oxford: Basil Blackwell, 1989.

Frith, Uta and Francesca Happé. “Autism: beyond ‘theory of mind.’” *Cognition*. Vol. 50. (1994): 115-132.

Grant, Cathy M., Kevin J. Riggs, and Jill Boucher. “Counterfactual and Mental State Reasoning in Children with Autism.” *Journal of Autism and Developmental Disorders*. Vol. 34, No. 2. (2004): 177-188.

Green, Joshua. “The Secret Joke of Kant’s Soul.” In *Moral Psychology Volume 3. The Neuroscience of Morality: Emotion, Brain Disorders, and Development.* 35-79. Edited by Walter Sinnott-Armstrong. Massachussetts: MIT Press, 2008.

Haidt, Jonathan. “The Emotional Dog and Its Rational Tail: A Social-Intuitionist Approach to Moral Judgment.” *Psychological Review*. Vol. 108, No. 4. (2001): 814-834.

Harris, Paul, Tim German, and Patrick Mills. “Children’s use of counterfactual thinking in causal reasoning.” *Cognition*. Vol. 61. (1996): 233-259.

Hill, Elisabeth L. “Evaluating the theory of executive dysfunction in autism.” *Developmental Review*. Vol. 24. (2004): 189-233.

Koster-Hale, Jorie, Rebecca Saxe, James Dungan, and Liane L. Young. "Decoding moral judgments from neural representations of intentions." *Proceedings of the National Academy of Sciences of the United States of America*. Vol. 110, No. 14. (2013): 5648-5653.

Leevers, Hilary J. and Paul L. Harris. “Counterfactual Syllogistic Reasoning in Normal 4-Year-Olds, Children with Learning Disabilities, and Children with Autism.” *Journal of Experimental Child Psychology*. Vol. 76. (2000): 64-87.

Leslie, Alan M., Ron Mallon, and Jennifer A. DiCorcia. "Transgressors, victims, and cry babies: Is basic moral judgment spared in autism?" *Social Neuroscience*. Vol. 1. (2006): 270-283.

Loveland, Katherine A., Deborah A. Pearson, Belgin Tunali-Kotoski, Juliana Ortegon, and M. Cullen Gibbs. "Judgments of Social Appropriateness by Children and Adolescents with Autism." *Journal of Autism and Developmental Disorders*. Vol. 31, No. 4. (2001): 367-376.

McGeer, Victoria. "The Varieties of Moral Agency: Lessons from Autism (and Psychopathy)." In *Moral Psychology Volume 3. The Neuroscience of Morality: Emotion, Brain Disorders, and Development.* 227-257. Edited by Walter Sinnott-Armstrong. Massachussetts: MIT Press, 2008.

McKenna, Michael. “Reasons Reactivity and Incompatibilist Intuitions.” *Philosophical Explorations*. Vol. 8, No. 2. (2005): 131-143.

Mele, Alfred. “Fischer and Ravizza on Moral Responsibility.” *Journal of Ethics.* Vol. 10, No. 3. (2006): 283-294.

Moran, Joseph M., Liane L. Young, Rebecca Saxe, Su Mei Lee, Daniel O'Young, Penelope L. Mavros, and John D. Gabrieli. "Impaired theory of mind for moral judgment in high-functioning autism." *Proceedings of the National Academy of Sciences of the United States of America*. Vol. 108, No. 7. (2011): 2688-2692.

Ozonoff, Sally, Bruce F. Pennington, and Sally J. Rogers. “Executive Function Deficits in High-Functioning Autistic Individuals: Relationship to Theory of Mind.” *Journal of Child Psychology and Psychiatry.* Vol. 32, No. 7. (1991): 1081-1105.

Peterson, Donald M. and Dermot M. Bowler. “Counterfactual Reasoning and False Belief Understanding in Children with Autism.” *Autism*. Vol. 4, No. 4. (2000): 391-405.

Rajendran, Gnanathusharan and Peter Mitchell. “Cognitive Theories of Autism.” *Developmental Review*. Vol. 27. (2007): 224-260.

Scott, Fiona J., Simon Baron-Cohen, and Alan Leslie. “If Pigs Could Fly: A Test of Counterfactual Reasoning and Pretense in Children with Autism.” *British Journal of Developmental Psychology*. Vol. 17. (1999): 349-362.

Shoemaker, David. “Psychopathy, Responsibility, and the Moral/Conventional Distinction.” *The Southern Journal of Philosophy*. Vol. 49, Spindel Supplement. (2011): 99-124.

Thomson, Judith Jarvis. “The Trolley Problem.” *The Yale Law Journal*. Vol. 94, No. 6. (1985): 1395-1415.

Todd, Patrick and Neal Tognazzini. “A Problem for Guidance Control.” *The Philosophical Quarterly*. Vol. 58, No. 233. (2008): 685-692.

Zalla, Tiziana, Luca Barlassina, Marine Buon, and Marion Leboyer. "Moral judgment in adults with autism spectrum disorders." *Cognition*. Vol. 121. (2011): 115-126.

1. I will be primarily focused here on their presentation of the theory in Fischer and Ravizza (1998) since this book has been the *locus classicus* for the vast majority of subsequent literature on their work and on reasons-responsive views more generally. However, their presentation of it there is, in many ways, a restatement and elaboration of some of their previous work. See, especially, Fischer (1987), Fischer and Ravizza (1991), Fischer (1994), and Fischer and Ravizza (1994) [↑](#footnote-ref-1)
2. Their view has a great deal of nuance that I am ignoring here since my focus will primarily be on their conception of moderate reasons-responsiveness as this is the feature that I wish to challenge in the next section. However, the focus on an actual sequence approach is central to the view (Fischer has defended this approach in a number of places, including Fischer (1982), Fischer (2007), and, more recently, Fischer (2015)) as is the suggestion that it is the agent’s mechanism (i.e. the process which leads to the action) and not the agent herself that must be reasons-responsive. [↑](#footnote-ref-2)
3. In subsequent work, Fischer refers to the view stated here as a “self-expression” account of moral responsibility. His central claim is that when an agent exercises guidance control in this moderately reasons-responsive manner she is expressing something about herself. More specifically, she is constructing a narrative of her own life. To this end, he writes, “what is expressed by an agent in acting is the meaning of the sentence of the book of his life.” (Fischer 1999, 290) It is this sort of unhindered self-expression, then, which gives value to our conception of moral responsibility. [↑](#footnote-ref-3)
4. See, for example, Baron-Cohen, Leslie, and Frith (1985) [↑](#footnote-ref-4)
5. See Ozonoff, Pennington, and Rogers (1991)For a review of this literature, see Hill (2004) [↑](#footnote-ref-5)
6. See Frith (1989); Frith and Happé (1994) [↑](#footnote-ref-6)
7. For example, the theory of mind hypothesis seems not to be able to explain the restrictive interests and repetitive behaviors that are characteristic of ASD while the executive function and weak central coherence hypotheses seem to have difficulty accounting for a number of the social features of the disorder. For a review of these theories and their relative strengths and weaknesses, see Rajendrand and Mitchell (2007). [↑](#footnote-ref-7)
8. [DELETED FOR BLIND REVIEW] [↑](#footnote-ref-8)
9. Similar results were reported in Scott, Baron-Cohen, and Leslie (1999). [↑](#footnote-ref-9)
10. In the standard false belief task (Baron-Cohen, Leslie, and Frith (1985)), participants are shown a story in which one puppet places a ball into a box and then leaves the room. While the first puppet is away, a second puppet comes and transfers the ball from the box into a basket. The second puppet leaves, the first puppet returns, and the participant is asked where the first puppet thinks the ball is. The aim of the task is to determine whether or not participants can identify that others may have false beliefs and predict behavior on the basis of these beliefs. [↑](#footnote-ref-10)
11. The following is an example of one such story: “This is Mary. Mary wants to find her kitten. Mary’s kitten is really in the bedroom. Mary thinks her kitten is in the kitchen. Where will Mary look for her kitten?” (179) [↑](#footnote-ref-11)
12. For additional evidence, see Begeer et al. (2009) [↑](#footnote-ref-12)
13. [DELETED FOR BLIND REVIEW] [↑](#footnote-ref-13)
14. Some of the material in this section draws on [DELETED FOR BLIND REVIEW] [↑](#footnote-ref-14)
15. See Blair (1996); Leslie, Mallon, and DiCorcia (2006) [↑](#footnote-ref-15)
16. See Shoemaker (2011) for a discussion of these. [↑](#footnote-ref-16)
17. See Moran, et. al. (2011); Koster-Hale et. al. (2013). [↑](#footnote-ref-17)
18. For more on this, see McGeer (2008). [↑](#footnote-ref-18)
19. [DELETED FOR BLIND REVIEW] [↑](#footnote-ref-19)
20. Fischer (2005) says this explicitly in response to an objection from Michael McKenna. [↑](#footnote-ref-20)
21. Patrick Todd and Neal Tognazzini raise a similar worry. More on this below. [↑](#footnote-ref-21)
22. Admittedly, there could be cases where the deficit is so deep and pervasive that the agent may not be regularly receptive to reasons. For example, we might imagine an agent who could *never* distinguish intentional harms from accidental harms and, as a result, becomes irate at those who accidentally harm others. Such an agent would seem to fall short of Fischer and Ravizza’s conditions for regular receptivity. However, this sort of case, I think, overstates what is shown by the empirical studies on ASD. These studies suggest that the ability to make these distinctions is *impaired* in ASD, not that it is altogether *absent*. I grant that such a case would not be a problem for Fischer and Ravizza’s view, but the empirical data suggest that the impairment (at least in cases of high-functioning agents) is much less severe. (For evidence that the impairment may not be as severe as some suppose, see Channon et al. (2010); Channon et al. (2011); Buon et al. (2013)). I am grateful to an anonymous referee for raising this worry. [↑](#footnote-ref-22)
23. See Haidt (2001); Green (2008); Cushman, Young, and Greene (2010) [↑](#footnote-ref-23)
24. See Crockett (2013); Cushman (2013); Cushman (2015); Dolan & Dayan (2013) [↑](#footnote-ref-24)
25. Mele uses the case of an agoraphobic agent while McKenna rests his argument on a hypothetical case of an agent that is reactive to one and only one reason. See, Mele (2006); McKenna (2005) [↑](#footnote-ref-25)
26. See McKenna (2005) for a discussion of these reasons. [↑](#footnote-ref-26)
27. They define this notion as follows: “An actually operative kind of mechanism is receptive to the actual reason if and only if there is a world in which the same mechanism operates and the same sufficient reason to do otherwise is present and the agent recognizes the sufficient reason to do otherwise.” (692). [↑](#footnote-ref-27)