Journal of the American Philosophical Association (2017) 446–464 © American Philosophical Association
DOI: 10.1017/apa.2017.37

Autism, Metacognition, and the Deep Self

ABSTRACT: *Many 'deep self' theories of moral responsibility characterize the deep self as necessarily requiring that an agent be able to reflect on her own cognitive states in various ways. In this paper, I argue that these metacognitive abilities are not actually a necessary feature of the deep self. In order to show this, I appeal to empirical evidence from research on autism spectrum disorders (ASD) that suggests that individuals with ASD have striking impairments in metacognitive abilities. I then argue that metacognitive conceptions of the deep self are implausible insofar as they fail to give a satisfactory account of the responsibility of persons with autism.*

KEYWORDS: moral responsibility, moral agency, deep self, autism spectrum disorder

Introduction

'Deep self' theories of 'moral responsibility claim, most generally, that an agent is responsible for her actions or attitudes in virtue of the fact that they are expressive of her deep, or real, self. The various deep self views are distinguished from one another, then, by the way in which they specify what an agent's real self consists in. For example, on Frankfurt's famous account the real self is grounded in a particular hierarchical relationship between an agent's desires, and insofar as these properly ordered desires are expressed in an agent's actions, the agent acts freely. Alternatively, Gary Watson has offered an account that appeals to the interaction between the agent's desires and her evaluative judgments in order to characterize the deep self. On his view, an agent acts freely to the extent that her motives and her values are aligned. What these views (and many others) share, however, is a reliance on some type of metacognitive reflection as a necessary feature of the deep self. I will argue that reflection of this sort is not necessary for an adequate conception of the deep self, and in order to show this I will appeal to empirical work on a group of individuals who seem to be profoundly impaired in this sort of reflection—namely, high functioning individuals with autism spectrum disorder (ASD)—but who nevertheless seem to be morally responsible agents. My aim is not to advance a particular conception of responsible agency but to show that one prominent and influential approach is ultimately unsuccessful.

I am grateful to David Shoemaker, Michael McKenna, Victoria McGeer, and Alison Denham for their generous comments and advice on earlier drafts of this paper. I would also like to thank Nathan Biebel, Christopher Boom, Jesse Hill, and Dan Tigard for their helpful comments.

CrossMark
https://doi.org/10.1017/apa.2017.37 Published online by Cambridge University Press

1. Frankfurt and Watson on the Deep Self

In his, 'Freedom of the Will and the Concept of a Person,' Frankfurt's central claim is that both freedom of the will and personhood are best understood by reference to the structure of an agent's desires. Such a view can make sense of the intuitive idea that part of what it means for an agent to act freely is simply for her to act in the way she wants to act, and Frankfurt offers a number of distinctions in order to shed light on just how an agent's desires are central to her being morally responsible for her actions.

He begins by making a distinction between an agent's will and an agent's second-order volitions. The will, according to Frankfurt, is simply an effective first-order desire (i.e., a desire that actually leads an agent to act). A second-order volition, on the other hand, is a desire for a particular first-order desire to be one's will. According to Frankfurt, whether or not an agent is morally responsible is to be determined by the hierarchical relationship between her effective first-order desires and her second-order volitions. If an agent's actions issue from a will that she wanted to have, then her will is free and those actions are *hers* in an important sense. Additionally, having second-order desires is a necessary condition for one's being a person, and those who lack these higher-order desires are, in Frankfurt's terms, wantons.

Frankfurt's picture of what it means for an agent to be morally responsible is an intuitively appealing one. It captures something striking about actions for which agents can be held responsible, namely, that they reflect certain important facts about the agent herself or about who the agent *really* is. However, Frankfurt's view is not without its problems. The most notable among these is the fact that there seems to be nothing particularly special about second-order volitions that would license our affording them such a central role in grounding such important concepts as personhood and responsibility or, as Gary Watson puts it, self-determination. That is, if second-order volitions are nothing more than desires, then why should we take them to have any special authority and what would stop us from requiring that they be desires that an agent wants to have in the same way that Frankfurt requires that the will be a desire that the agent wants to have? There is nothing in Frankfurt's picture, so far, that allows us to avoid such an infinitely ascending hierarchy of desires. Sensitive to this issue, Frankfurt attempts to stem the worry about a regress by invoking the notion of identification. Thus, he writes,

> It is possible, however, to terminate such a series of acts without cutting it off arbitrarily. When a person identifies himself *decisively* with one of his first-order desires, this commitment 'resounds' throughout the potentially endless array of higher orders. Consider a person who, without reservation or conflict, wants to be motivated by the desire to concentrate on his work. The fact that his second-order volition to be moved by this desire is a decisive one means that there is no room for questions concerning the pertinence of desires or volitions of higher orders. (Frankfurt 1971: 21)

https://doi.org/10.1017/apa.2017.37 Published online by Cambridge University Press

That is, Frankfurt thinks it possible to cut off nonarbitrarily any potential regress by *identifying* with the desire that constitutes one's will.

This portion of Frankfurt's original view has been the subject of forceful criticism by Watson (1975) who denies that Frankfurt's proposed source of authority for second-order volitions can do the work that Frankfurt wants it to do. Watson's central claim here is just that Frankfurt's proposed solution on the basis of identification with a particular desire simply cannot avoid the charge of arbitrariness. As Watson puts it,

> We wanted to know what prevents wantonness with regard to one's higher-order volitions. What gives these volitions any special relation to 'oneself'? It is unhelpful to answer that one makes a 'decisive commitment', where this just means that an interminable ascent to higher orders is not going to be permitted. This *is* arbitrary. (Watson 1975: 349)

In responding to this criticism, Frankfurt clarifies his view by claiming that the identification needed in order to assign second-order volitions the authority they need comes in the way of a wholehearted decision by the agent to constitute herself in accordance with the desire in question (Frankfurt 1987). Central to Frankfurt's response is the claim that when one makes a *decisive* commitment to a certain desire, he does so 'in the belief that no further accurate inquiry would require him to change his mind' (1987: 169). Frankfurt continues,

> It is therefore pointless to pursue the inquiry any further. . . . [A] person can without arbitrariness terminate a potentially endless sequence of evaluations when he finds that there is no disturbing conflict, either between results already obtained or between a result already obtained and one he might reasonably expect to obtain if the sequence were to continue. Terminating the sequence at that point—the point at which there is no conflict or doubt—is not arbitrary'. (1987: 169)

That is, when an agent makes a decisive commitment to a particular desire, she is deciding to endorse that desire and is thereby creating herself in a distinctive way (Frankfurt uses the language 'making up one's mind' here to convey the notion that the agent is literally configuring her mind in a certain fashion). When such a decision is made wholeheartedly—that is, when an agent, in deciding, establishes certain patterns of response or certain preferences in her mental economy—identification becomes an act sufficiently authoritative as to give the agent's second-order volitions the quality that Frankfurt's account suggests and to overcome the charge of arbitrariness leveled by Watson. (In later work, Frankfurt characterizes wholeheartedness not as a function of identification but as a function of a kind of self-satisfaction; see Frankfurt [1992].)

In his response to Frankfurt, Watson denies that Frankfurt's conception of identification can ground an agent's will and suggests, instead, that an agent's values

https://doi.org/10.1017/apa.2017.37 Published online by Cambridge University Press

and their connection to her desires may be a better candidate for identifying the real self. Given his objections, canvassed above, to the hierarchical nature of Frankfurt's view, Watson offers a competing, dual-system view in which an agent is seen as being free in the sense required for moral responsibility only in cases where what the agent values and what the agent is moved to do are one and the same. Rather than understanding responsibility as stemming from a properly ordered hierarchical relationship between desires, Watson conceives of it as stemming from the fact that an agent's motivational system and her evaluative system are acting in harmony with one another. As he puts it, 'If there are sources of motivation independent of the agent's values, then it is possible that sometimes he is motivated to do things he does not deem worth doing. This possibility is the basis for the principal problem of free action: a person may be obstructed by his own will'" (Watson 1975: 345).

In characterizing these separate systems Watson relies on a broadly Platonic conception of practical reasoning in which reason both determines what things have value and provides motivation for an agent to act. Thus, he writes,

> Since the notion of value is tied to (cannot be understood independently of) those of the good and worthy, it is one thing to value (think good) a state of affairs and another to desire that it obtain. However, to think a thing good is at the same time to desire it (or its promotion). Reason is thus an original spring of action . . . values provide *reasons* for action. The contrast is with desires, whose objects may not be thought good and which are thus, in a natural sense, blind or irrational. Desires are mute on the question of what is good. (Watson 1975: 340)

This distinction provides the basis for Watson's central claim because it allows for conflict between one's 'valuational system' and one's 'motivational system,' between the desires of reason and desires that do not reflect one's values.

Given that Watson rejects the hierarchical structure employed by Frankfurt, it may seem initially that his view is not essentially metacognitive. That is, if the point is just that sometimes agents are motivated by desires that accord with their values and sometimes by desires that do not, then there would seem not to be any essential metacognitive element in the view at all. However, this is not quite right, and we can see why by attending to the ways in which Watson suggests that conflicts between desires and values can arise, of which there are two.

First, Watson argues that this conflict can arise by way of an agent's experiencing an 'estranged desire,' a desire that is entirely divorced from the agent's values. As instances of estranged desires Watson offers examples of a mother who suddenly has a desire to drown her bawling infant in the bathtub or a man who believes that his sexual urges are the work of the devil. In each of these cases it is just false that the agent values what is desired—clearly the mother does not value drowning her baby, for example. In order for an agent to be estranged from her desires (and, thus, in order for the problem of free agency to exist) it must be the case that the agent can reflect on her desires *through the filter of her values*. Without the capacity to reflect on the desire in question, no sense of estrangement would arise.

https://doi.org/10.1017/apa.2017.37 Published online by Cambridge University Press

The second way in which desires and values may come into conflict, Watson suggests, arises from the possibility that in some cases what one most strongly desires may not be what one most values. Here the idea is just that it may be possible that some actions are driven by desires that are separate from one's values and that may be strong enough to override the desires of reason. Importantly, a metacognitive capacity is implicated in this case as well. For Watson holds that reflecting on these desires is a crucial part of the story because some purely appetitive desires are desires that agents value having. For example, the desires for food or sex are desires that most of us value having, that is, we would consider it a loss if we ceased to desire these things. Conflict arises, then, only when we reflect on the desires in question and judge that they ought not to be as strong as they are. Thus, in this second case, as in the first, it is crucial, on Watson's picture, that an agent be able to reflect on her desires. An agent who lacked this capacity would be an agent for whom there simply is no problem of free action.

Here, then, as in Frankfurt's view, we find a necessary condition of self-reflection in the exercise of free agency. In order to be free, the agent must be able to reflect on her desires in light of her evaluative judgments, and this opens up the possibility of her being obstructed by her own will. This requirement of self-reflection, however, makes both Frankfurt and Watson vulnerable to an important objection from the empirical literature on ASD.

2. Metacognition in ASD

The reason that evidence regarding autism is helpful in assessing views like Frankfurt's and Watson's is that the very capacity that they take to be central to a working conception of the self—metacognition that is, the ability to reflect on or cognize about cognitive states—seems to be seriously impaired in ASD. Yet, there is a strong intuition that high-functioning autistic persons are morally responsible for, at least, a wide-ranging set of actions (given that autism is a spectrum disorder and can present in very different ways, reference to autistic persons, for the purposes of this paper, should be understood as referring to those on the high-functioning end of the autism spectrum).

Evidence of a metacognitive impairment in ASD comes from two central sources. The first is the ample evidence suggesting that individuals with autism have difficulty understanding the representational nature of belief. Baron-Cohen, Leslie, and Frith (1985), for example, famously showed that individuals with ASD perform well below neurotypical levels on traditional false belief tasks that are intended to test the subject's ability to attribute a false belief to another person. Moreover, studies have shown that autistic individuals show signs of a deficit in understanding their own beliefs as well. Such a failure may be plausibly interpreted as a failure to understand the representational nature of belief, and other evidence has been gathered that supports this (see, for example, Baron-Cohen et al. [1999]; Gallagher and Frith [2003]; Saxe and Kanwisher [2003]; and Koster-Hale et al. [2013]; for reviews of this literature, see Frith, Morton, and Leslie [1991] or Rajendran and Mitchell [2007]). This evidence, collectively, underwrites the influential

https://doi.org/10.1017/apa.2017.37 Published online by Cambridge University Press

'theory-of-mind' hypothesis regarding the nature of autism, which holds that 'individuals with autism fail to "impute mental states to themselves and others" and that this deficit manifests as inability to mentalize, or failure to take into account others' mental states' (Rajendran and Mitchell 2007: 226, quoting Premack and Woodruff 1978). An impaired ability to understand the representational nature of mental states would seem to entail impaired metacognition.

A second source of evidence comes from similar studies regarding desire understanding in ASD. Given that reflecting on one's desires is crucial for both of the deep self views discussed above, it will be especially striking if desire understanding is impaired in ASD, and, indeed, this seems to be borne out in the empirical literature. While research that specifically targets desire in ASD is comparatively sparse, there is nevertheless some evidence available suggesting that individuals with ASD are less able to understand desire than are neurotypical individuals. More specifically, this evidence seems to show that autistic persons are less able to understand the representational nature of desire and, instead, seem to understand desirability as a property of the objects.

Consider, for example, a study conducted by Phillips, Baron-Cohen, and Rutter (1995) that found that children with ASD showed signs of severe impairment relative to neurotypical and mentally handicapped participants in both the ability to recognize when the desires of others were unsatisfied and the ability to understand the representational content of desires. In the first experiment of this study, participants were presented with vignettes about actors whose desires were satisfied and others whose desires went unsatisfied. In half of the stories the desires of the actor were made explicit, and in the other half participants were required to infer the desire prior to judging satisfaction. The researchers found that individuals with ASD performed at near control levels on the explicit tasks but showed signs of significant impairment on the implicit tasks (54% of autistic participants passed the explicit task while only 13% passed the implicit task whereas control participants passed the tasks at rates of 77% and 67% respectively). What is suggested by this study is that individuals with ASD are capable of matching goals with their corresponding successful outcomes, but that they seem much less able to infer the desires of others when the goal or desire is not provided explicitly.

In a second experiment, the authors tested the ability of participants to recognize when another person's desire has changed. The goal of this experiment was to test the ability of participants to recognize the representational nature of desire. As they put it,

> The same entity (object, event, situation) can appear desirable to one person, while at the same time being undesirable to another . . . to understand this fact requires a concept of desire that includes its personal, subjective nature. One needs to understand that desirability is not a property of the object, but of the person's mental representation of the object. (Phillips, Baron-Cohen, and Rutter 1995: 160–61)

In order to test this ability, participants were presented with vignettes in which characters are described as beginning to perform some action, changing their minds,

https://doi.org/10.1017/apa.2017.37 Published online by Cambridge University Press

and then performing a different action, and participants were asked to identify the desire of the character before he or she changed his/her mind. The results showed that typically developed children and children with mental handicap performed at near-ceiling levels while children with ASD performed substantially worse. The authors summarized the results as follows:

> In this experiment, it was not the satisfaction conditions of desire that needed to be understood . . . it required participants to realize that desires exist at the mental level, representing aspects of the world as desirable or not desirable. It would appear that children with autism are impaired in the ability to understand this aspect of desire, compared with children with mental handicap and normal children. Although there was some indication that the task is less difficult than understanding false belief, an autism-specific deficit was still apparent. (Phillips, Baron-Cohen, and Rutter 1995: 165)

Based on this research, it would appear that individuals with ASD do, indeed, show signs of impairment in the ability to understand the nature of desires as states that occur in the mind.

Similar results were obtained in a more recent study conducted by Broekhof and colleagues. (2015) that tested the ability of participants to recognize the subjective nature of desire. In this study, participants were given stories containing information about a character's preferences. In half of these stories, the character's preference was the same as the preference of the participant, and in the other half the preferences of the two were different. Participants were then asked questions about how the character in the story would behave on the basis of his or her desires. On tasks in which the desire of the character in the story matched the desire of the participant, participants with ASD performed at the same level as typically developed participants. However, on tasks where the preferences of the two were dissimilar, participants with ASD performed significantly worse than typically developed controls. What these results suggest, the authors claim, is that individuals with ASD seem to be impaired in their ability to recognize that the desires of others differ from their own. If this is correct, then it would support the argument from Phillips and colleagues that autistic individuals seem not to grasp the representational nature of desire. To put it another way, for autistic persons it seems to be the case that the property of being desirable attaches to the object itself rather than to one's representation of the object. On this conception, it would not be the case that an object is desirable *to me*. Rather, it would be the case that the object simply *is desirable*, and to have a desire is to be drawn to (or to have some typical phenomenological experience of wanting) the object in virtue of its having this property. To help avoid confusion, I will refer to this unique sense of desirability as 'desirability*' in the following discussion.

Given this evidence, it can be concluded that individuals with ASD show signs of impaired understanding of desire. It is important to note that the nature of the evidence with respect to desire is much the same as the evidence on belief

https://doi.org/10.1017/apa.2017.37 Published online by Cambridge University Press

understanding presented above, which, as we saw, posits an atypical understanding not only of the beliefs of others but of one's own beliefs as well. While the studies cited here do not take up the question of whether individuals with ASD understand their own desires in an ordinary way, it can be reasonably inferred that they do not. Given the similarity between these experimental methods and those in studies of belief understanding, it is likely that individuals with ASD are impaired in their understanding of desires in general, including their own desires, and this should not be especially surprising given the representational similarities between desire and belief more generally. With this evidence presented, we can now turn to the issue before us: how, if at all, does this impairment of metacognition help to shed light on the deep self views described in section 1?

3. The Challenge for Metacognitive Theories

Obviously, individuals with ASD have desires. Just as obviously, those desires can be effective in issuing in action. That is, individuals with ASD clearly have a will in the Frankfurtian sense. The important question for Frankfurt's view is whether individuals with ASD can have second-order volitions. Given the evidence just presented, it is not clear what the answer to this question would be. No research has been conducted, to my knowledge, on the presence of second-order volitions in ASD, and thus our best hope for answering this question is to do so speculatively on the basis of desire understanding in ASD more generally. It seems to me that there may be good reason to doubt that such second-order volitions are present. The general conclusion to draw from the evidence regarding desire understanding is that individuals with ASD seem impaired in their ability to understand that desires are mental representations.

The reason that this is problematic for Frankfurt lies in the fact that his view requires a substantial degree of self-reflection on one's desires *as the desires of one's self*. The crucial feature of the wanton is that he simply does not care to reflect on his first-order desires and so does not have second-order volitions as a result. The problem that the evidence on desire understanding in ASD poses comes out of this issue of self-reflection as well. However, while the wanton fails to reflect on his first-order desires simply because he does not care what they are or what they will be in the future, my suggestion is that individuals with ASD do not engage in this sort of self-reflection because they do not have a sufficient grasp of desire as something that is fundamentally subjective. To reflect on desire in ASD, then, is to reflect on the desirability* of some object, but given this conception of desire, this is not to reflect on one's *will*. It is the latter sort of reflection that Frankfurt's account requires.

Perhaps an example will be helpful here. Suppose that Tom is a high-functioning autistic person who desires to smoke a cigarette. On the basis of this desire, he lights up a smoke and puffs away. His will is to smoke the cigarette. Now, the important question, on Frankfurt's view, is whether or not Tom wants this to be his will. How will Tom answer this question? If the studies cited above give a correct picture of Tom's understanding of desire, then it is hard to see how he could make sense of

https://doi.org/10.1017/apa.2017.37 Published online by Cambridge University Press

this. It is possible that he might bemoan the fact that desirability* is a property of cigarettes. That is, he might desire that cigarettes not be desirable*, but, crucially, this is not a desire about his will. Rather, it is a desire about cigarettes. It is on a par with desiring that cigarettes not be unhealthy or desiring that they not smell bad or the like. As such, this desire is not a second-order volition at all. The key problem for Frankfurt's account that the evidence on desire in ASD presents, then, is this: given that desires are taken to be the desirability* of their objects, any second-order volition in ASD that purports to play the role that Frankfurt's view needs it to play will fail to do so because such desires would turn out not to be about one's will at all.

If this speculative conclusion regarding the absence of second-order volitions in ASD is correct, then how would autistic individuals be classified on Frankfurt's account? It seems as though his view would entail that individuals who have impairments such as these in desire understanding are, in fact, wantons because on his view, a wanton is just an individual who lacks second-order volitions. However, this is an extraordinarily counterintuitive implication of the view because it entails not only that many individuals with ASD are not responsible agents but also that they are not even persons. Such an implication can easily be falsified by anyone who has spent any amount of time conversing with a high-functioning autistic person. Thus, if Frankfurt is committed to this view of individuals with ASD, then this seems a good reason to reject the view.

However, there may be a response open to Frankfurt here. If it is correct that someone like Tom in the example above could have the desire that desirability* not be a property of cigarettes, then perhaps a desire of this sort could play the functional role of second-order volitions and prevent wantonness in individuals with ASD. After all, Tom does appear to be relevantly different from Frankfurt's case of the willing addict. The fact that Tom wants cigarettes not to be desirable* sets him apart, in some way, from the person who is wholly delighted that they have this property. If desires about the desirability* of particular objects play a functionally equivalent role to second-order volitions in Frankfurt's view, then perhaps he can say that individuals with ASD can be morally responsible agents by virtue of having *these* desires even in the absence of second-order volitions. This response might be sufficient to show that individuals with ASD have something like second-order volitions, but it will not succeed in saving Frankfurt's view from the broader challenge from ASD. If desires such as these are to play the role of second-order volitions, it is not enough for them to have the right sort of content; they must also occupy the authoritative position that Frankfurt assigns to second-order volitions.

On Frankfurt's view, second-order volitions attain their authoritative role when the agent identifies wholeheartedly with them or when she is satisfied with them. This brings us to the second question posed above: can individuals with ASD engage in this form of identification? It is difficult to see how they could. Recall that wholehearted identification involves a decisive commitment to constitute oneself according to the desire in question, and it is unlikely that one could constitute oneself according to a first-order desire like the ones I have just described. That is, perhaps Tom could arrive at the conclusion that no further accurate deliberation

https://doi.org/10.1017/apa.2017.37 Published online by Cambridge University Press

would lead him to give up his desire that cigarettes not be desirable*, but there seems to be no plausible story to tell about how this sort of commitment to a desire like this could *constitute* Tom in any way since Tom (or some essential feature of Tom) is not the object of the desire. More important, Tom's desire for cigarettes and his desire that cigarettes not be desirable* do not seem to pose any sort of agential conflict in the way that the first-order desire and the second-order volition of Frankfurt's unwilling addict seem to do. The unwilling addict wants his will to be something other than it is, and it is the fact that it is *his will* that creates puzzles about his responsibility. In Tom's case, however, no such conflict arises. His desire to smoke is no more incompatible with his desire that cigarettes not be desirable* than it would be with a desire that cigarettes be less expensive, say.

The problem here, then, is that desires like Tom's are not such that they lend themselves to either identification or satisfaction in Frankfurt's sense. Thus, the evidence on desire understanding in ASD casts doubt on the claim that desires could ground the real self of autistic individuals. If real self views are to meet the challenge of correctly characterizing autistic agents, we will need to look to some other agential feature to do the grounding work.

A similar problem, it seems to me, arises for Watson's account. If individuals with ASD lack cognitive access to desires as mind-dependent representations, then the sort of self-reflection Watson takes to be crucial is ruled out. In other words, given the impairments of desire understanding outlined above, reflecting on desire for individuals with ASD would amount to reflecting on a certain property of some object or action rather than reflecting on oneself. If this is correct, then it would make little difference whether the desire conformed to one's values. Instead, the object would be seen as something that is desirable regardless of one's attitude toward it. The same issue posed for Frankfurt, therefore, comes up for Watson as well. Reflection on desire in ASD is simply not, or so it would seem, *self*-reflection. To put the point another way, if one lacks an adequate understanding of one's mental states, then one will not be able to reflect on those states in the way that Watson's and Frankfurt's views require. Whereas the evidence on desire understanding in ASD led Frankfurt to the implausible conclusion that autistic individuals are not persons, it seems to lead Watson to the absurd conclusion that autistic persons *never* act freely.

The evidence on metacognition in ASD, then, casts some doubt on the plausibility of both Frankfurt's and Watson's general accounts of the real self. However, the objection that I have posed against Watson raises an important problem that requires a solution. The account of the interaction between desires and values in ASD that I have been defending entails that the desires of individuals with ASD are not regulated by their values because such regulation would require a robust self-reflective ability. If this is true, though, the issue of estranged desires like those Watson appeals to becomes especially pressing. For if these sorts of desires are not regulated or suppressed through some self-reflective process, then we should expect to see autistic persons acting on desires like those Watson cites. That is, we should expect, for example, to see autistic mothers drowning their bawling infants when the desire strikes them. Clearly, though, this is not the case. Autistic persons, though

https://doi.org/10.1017/apa.2017.37 Published online by Cambridge University Press

impaired in social interaction, generally display no tendencies toward violent or antisocial behavior. Clearly these sorts of desires, if they exist, are being successfully regulated or suppressed, and an explanation of how this is accomplished is needed.

4. Toward a Conception of the Autistic Deep Self

An account of how autistic persons succeed in inhibiting aberrant desires like those Watson discusses may bring us close to an alternative account of the deep self, one that would capture what it is for an action truly to belong to a person without appealing to metacognition. One way of getting at an explanation of the ability to regulate or suppress these sorts of desires is by examining the evidence on inhibitory control in individuals with ASD.

While many studies have shown that individuals with ASD have deficits in certain executive functions (see Ozonoff, Pennington, and Rogers [1991]; Rajendran and Mitchell [2007]; for an authoritative review of studies of executive function in ASD see Hill [2004a]), inhibitory control in the disorder is not well understood. However, the most plausible picture seems to be that autistic persons demonstrate impairments in their ability to inhibit *prepotent* responses (i.e., responses that the individual is strongly predisposed to have) but that they perform at control levels on all other inhibitory tasks. (Hill 2004a, 2004b) In other words, when confronted with decisions regarding how to act, individuals with ASD are generally able to exercise inhibitory control unless doing so requires inhibiting a response to which they are strongly predisposed. This evidence suggests one immediate explanation for why individuals with ASD do not act on desires like those that Watson describes, namely, that the responses to such desires (e.g., drowning one's infant) are not prepotent responses, and therefore, they do not override inhibitory control. Of course, this only pushes the problem back a step as we still need some explanation for why these responses are not prepotent, or, better, an explanation for why other responses *are* prepotent. The most natural and plausible explanation arises from the intact capacity for affective empathy in autistic persons as well as from the somewhat atypical nature of moral judgment in ASD. It is likely that these produce strong prepotent responses such that aberrant desires can be suppressed.

Despite what many have suggested (see Kennett [2002] and, to a lesser extent, Shoemaker [2015b]), affective empathy is largely unaffected in autistic persons. This is often disguised by the presence of impairments in the ability to understand the minds of other people (i.e., to interpret intentions, desires, and beliefs on the basis of actions), but the basic affective features of empathy remain intact for autistic persons. For example, people with autism tend to experience emotional contagion and, thus, are able to be moved by the emotions of others. Moreover, they tend to experience distress in response to others' negative emotions at normal levels, and they experience emotional concern for the well-being of others. (The emotional capacities of persons with autism are discussed at length in Stout [2016b] and Stout [2017]. See also Yirmiya [1992]; Blair [1999]; Rogers et al. [2007]; Dziobek et al. [2008]; Hirvelä and Helkama [2011].)

https://doi.org/10.1017/apa.2017.37 Published online by Cambridge University Press

As a first pass, perhaps we can say that individuals with ASD are able to regulate desires according to the extent to which those desires conform to standing emotional dispositions, specifically, in the moral case, dispositions characterized by prosocial attitudes toward others. Given that autistic persons possess intact emotional empathy abilities, it is likely that they, like most moral agents, have strong prosocial prepotent responses in the majority of circumstances (though there is some evidence that their ability to felicitously navigate social situations is compromised). Indeed, this seems to be what is borne out in the literature on moral judgment in ASD (see Blair 1996; Leslie, Mallon, and DiCorcia 2006; Stout 2016c). In short, autistic people are clearly able to engage in caring relationships, and they are just as clearly able to feel a strong emotional concern for others. Therefore, it is likely these emotional dispositions result in prepotent responses that would be sufficient to suppress rogue desires like the ones Watson is concerned with.

A second explanation for why autistic persons are able to inhibit desires such as these may come out of the atypical nature of moral judgments in ASD. More specifically, I have argued elsewhere (Stout 2016a) that individuals with ASD rely more heavily than do neurotypical people on model-free judgments due to the fact that these do not require the use of models that must be supported by counterfactual abilities. Here I am relying on a dual-system account of moral judgment that holds that judgments occur according to model-based and model-free processes (this sort of account has been compellingly developed by Fiery Cushman [2013], among others). Model-based judgments require the agent to build a mental model (like a decision tree of sorts) of the world and then to trace the outcome of a given judgment on the model. Model-free judgments, on the other hand, are automatic judgments made in response to certain stimuli that have been assigned either positive or negative value by the agent due to habituation over time as a result of feedback received in response to previous, similar judgments or behaviors. This fact is important because it allows judgments to proceed independently of any counterfactual or self-reflective considerations regarding desire satisfaction. That individuals with ASD rely more heavily on model-free judgments meshes well with accounts of 'compensatory strategies' developed by autistic persons. Thus, it could be the case that, due to their habituation over time, model-free judgments produce strong prepotent responses that would be sufficient to override anomalous desires.

There are, then, two plausible candidate features of the psychology of ASD that allow us to avoid the problematic implication of my objection to Watson, and the crucial feature of each of these is that neither requires that the agent be able to reflect on her desires in order for the action to be hers. Instead, all that would be required is that the desire be in line with the agent's standing emotional dispositions or with her habituated patterns of judgment. Notice, however, that what this gives us is the basis for an alternate conception of the real self, one that can accommodate the empirical data on ASD in a way that the views considered here could not. Applying this to Watson's case of the mother who desires to drown her bawling infant, we can see clearly that her desire does not align with her standing emotional dispositions (e.g., to care for her child, to be concerned with her child's well-being, etc.), and presumably, it does not match up with her pattern of model-free judgment. However, had the desire been such that it could not have been suppressed, even by

https://doi.org/10.1017/apa.2017.37 Published online by Cambridge University Press

these strong prepotent responses, then it seems plausible to say that she would not have acted freely had she acted on that desire.

This conception of the real self is similar to, though also importantly different from, a recent account proposed by David Shoemaker (2015a, 2015b). Shoemaker locates the real self in an agent's cares and commitments such that an agent is responsible for an attitude if it is 'causally dependent on, and its content is harmonious with, at least one of the agent's cares, commitments, or care-commitment clusters' (Shoemaker 2015b: 59). For Shoemaker, cares are 'dispositions to respond emotionally in sync with the fortunes of the cared-for object' (2015b: 51), and commitments are simply the values that are demonstrated by the sum total of the agent's evaluative judgments and, as a result, make up the agent's evaluative stance. Both cares and commitments, on Shoemaker's view, are indicators of what *matters to* an agent and thus mark something important about the deep self of the agent.[1]

Clearly, then, Shoemaker's appeal to cares corresponds closely to the emotional dispositions that I have been discussing here. However, there is an important difference between an agent's commitments, as he conceives them, and the pattern of judgments to which I have appealed. The agent's pattern of judgments is broader, it seems to me, than Shoemaker's notion of commitments. What I have in mind is simply the judgments that the agent makes over time under relevantly similar conditions. This pattern need not reveal any commitments about the agent's conception of the good and thus does not constitute an evaluative stance the agent takes. However, it does reveal something important and genuine about the way in which the agent typically governs herself across a wide range of circumstances. If an action or attitude is in line with this pattern, then it can be reasonably attributed to some feature of her psychology that is genuinely hers. Another crucial difference between the agent's patterns of judgment and her commitments, as Shoemaker conceives them, is that the agent's pattern of model-free judgments is modulated in important ways by her emotions. Shoemaker draws on Watson's view in describing commitments, and accordingly these turn out to be fundamentally rational. Importantly, though, model-free judgments are simply judgments that have been habituated via positive and negative feedback, and a central part of such feedback is the emotional response it occasions in the agent who receives it. One important upshot of the preliminary view I am offering here is that the emotions play a necessary role in an action or attitude being attributed to the agent, whereas on Shoemaker's account this need not be the case.

Admittedly, the notion of the real self just presented is inchoate, and it remains to be seen whether it can be generalized beyond the case of autism such that it can constitute a stand-alone account of the real self. Moreover, I do not wish to endorse this view as a general account of moral responsibility. I am simply suggesting that if a plausible account of the real self that can accommodate the evidence from autism

1 This is Shoemaker's most recent view of the deep self, and it is importantly different from the view that he presents in his influential (2003) paper. There, he presented a view of the deep self which appealed to cares exclusively (understood as fundamentally emotional). Since his more recent work reflects a modification of his earlier view, that work will be my focus here.

https://doi.org/10.1017/apa.2017.37 Published online by Cambridge University Press

is to be offered, it seems to me that it must draw on the features I have discussed here. The agent's emotional dispositions along with her established patterns of model-free judgment seem to be the only agential features that could support strong enough prepotent responses to preclude the agent from acting on aberrant desires such as those described by Watson.

5. Objections and Replies

Before closing, I will take time to consider two important objections to the arguments of this paper. First, as I noted in the introduction, there is a deep methodological objection lurking behind the arguments presented here. For the metacognitive theorists might simply shrug and claim that everything that I have said so far gives us as much reason to think that individuals with ASD are *not* responsible as it does to think that the theories discussed are mistaken. Perhaps all I have done here is point to some interesting reasons to think that those with ASD are not among the class of responsible agents. This is a difficult charge to dismiss, but there is good reason to think that the burden of proof is on those who suggest that the order of explanation is from theory to cases rather than the other direction. To show this, it will be helpful to say something more about the genesis of this theoretical approach and the aims that motivate real self theories.

To begin with, each of the theories discussed so far was presented subsequent to Frankfurt's famous denial of the claim that freedom and moral responsibility require the ability to do otherwise (see Frankfurt 1969). Because of this, the real self theorists can reasonably be seen as trying to identify what free action and responsibility *do* require if not this ability, and each of them points to some feature of the agent's psychology as an answer. Then, having identified some crucial psychological element, each of them uses as evidence in favor of the element in question the fact that it can give plausible explanations for a number of intuitive cases. Frankfurt's hierarchical view thus gains plausibility from its ability to handle the cases of the addicts discussed above, and Watson's view gives a helpful explanation of cases of what he calls 'estranged desire'— cases that are supposed to show us something important about the connection between desires and values. Each of these views proposes necessary and sufficient conditions for moral responsibility and then justifies these conditions primarily by their ability to give plausible accounts of the cases presented.

What I have suggested so far is this: individuals with ASD pose problems for these necessary and sufficient conditions in that they seem to fail to meet the necessary conditions on Frankfurt's and Watson's view yet are plausibly still responsible agents. Now, the methodological objection to what I have been doing in this paper says that, given these individuals' failure to satisfy these conditions in the right sort of way, we ought to hold on to our theory and simply conclude that individuals with ASD are not responsible rather than rejecting the theory. However, it seems to me that the theorists discussed here have more argumentative work to do before this response is licensed. Specifically, they need to tell us why we should think that imaginary cases of willing addicts and estranged desires should be more

https://doi.org/10.1017/apa.2017.37 Published online by Cambridge University Press

authoritative in informing our conception of moral responsibility than are real-world cases of autistic persons. Given the prevalence of autistic agents in the world and the strong intuition (which I have defended at some length in Stout [2016b]) that high-functioning autistic people are capable of responsible agency across a great variety of circumstances, it seems to me that the inability of a theory to give us a plausible account of them should count at least as heavily *against* a view as its ability to handle the sorts of cases discussed here count in its favor. Moreover, the argumentative burden is on those who deny this to tell us why.

The second objection that must be addressed is similar to the first. It is as follows: the empirical evidence regarding metacognition in individuals with ASD shows impairment in metacognitive abilities, but it does not show that those abilities are entirely absent. Therefore, it might be the case that autistic persons are responsible according to real self views but that this responsibility is mitigated in certain ways. If this objection is successful, then metacognitive real self views are at least partially vindicated in that they are not committed to the strong claim that those with ASD are not persons or are not responsible agents at all. This is an important challenge, but it can be resisted in at least two ways. First, one might deny the empirical claim that individuals with ASD are capable of metacognition some of the time. A second strategy would be to accept the empirical claim but to show that metacognitive real self theories prove problematic even on the more modest view in which these individuals' responsibility is mitigated rather than nullified. I will consider both of these strategies now.

In order to advance the first strategy, one might appeal to the use of elaborate compensatory strategies that individuals with ASD often employ. That is, one might claim that the reason that some autistic persons are able to perform as well as neurotypical participants on tests of metacognition is not that they are equally able to engage in metacognition but that they have learned over time how to compensate for their inability to do so by using other means of predicting behavior or of engaging in appropriate behavior themselves. While compensation of this sort does occur in some circumstances, it seems unlikely that it can explain every case of successful metacognition in persons with ASD. Additionally, while this response might offer a speculative account of the experimental data, it does not do much to account for the commonsense data in favor of some metacognitive ability that one gains simply by interacting with those with ASD. Therefore, this first response can only go so far in holding off the objection in question.

However, the second strategy for response seems to me to be considerably more powerful. While it is true that autistic persons are sometimes capable of metacognition, the impairments discussed above are not minor or inconsequential. Rather, they are well-documented, and they seem to be pervasive and behaviorally significant even in high-functioning autistic persons. This fact is problematic for metacognitive theories in two ways. First, to use Frankfurt as an example, it means that there must be some explanation provided for why individuals with ASD should be thought of as having substantially diminished personhood despite the strong intuition that high-functioning autistic persons are, indeed, persons in the same sense as the rest of us are. In other words, the methodological burden of proof would still be on the metacognitive theorist even if it is a lighter burden to bear. Second,

https://doi.org/10.1017/apa.2017.37 Published online by Cambridge University Press

and more important, insisting on active metacognition in particular circumstances will undoubtedly yield false negatives in our responsibility ascriptions for particular actions performed by individuals with ASD. Suppose, for example, that a high-functioning autistic person performs a morally bad action, say, breaking a promise. She understands that it is morally bad to break one's promises without justification, and she has an established pattern of model-free judgments that reflect this. However, on a particular occasion she decides she simply cares more about (or, she is more emotionally disposed to) staying home and reading a book than she does about keeping her promise to meet a friend for lunch. No metacognitive processes take place in this case, and yet surely her friend has grounds to blame her for breaking the promise. On the metacognitive views, however, this cannot be true since she was not able to reflect on the desire that led her to act in the way that she did. A vast array of mundane cases like this one could be devised with the same effect. If, in addition to the sorts of stable psychological features that I have proposed here, a further, reflective condition is necessary for responsibility, then autistic agents will not be responsible for a wide range of ordinary actions. While this is an easier bullet to bite than the claim that autistic persons are not responsible *at all*, it is still a deeply problematic result for metacognitive theories of the real self.

6. Conclusion

My goal here has been to scrutinize metacognitive theories of the deep self, that is, theories that claim that the ability to reflect on certain mental states is necessary for moral responsibility. I have presented the two most influential approaches to characterizing the real self and argued that each of these approaches seems to imply, implausibly, that individuals with ASD are not moral agents (or, in Frankfurt's case, even persons). Therefore, each of them fails to give an adequate account of the moral responsibility of individuals with ASD. I have tried to offer an account of what the real self in ASD, if there is one, might consist in, and I have appealed to broadly unreflective agential features in doing so.

In this paper, I have focused solely on bringing out the problems faced by Frankfurt and Watson in light of the evidence from autism research. However, in closing, I would like to advance some additional provisional challenges that this evidence might present, and I will do so by beginning with the rather mundane observation that lots of influential philosophical theories assign a great deal of importance to metacognition. As a result, insights from the literature on metacognition in ASD could potentially be seen as throwing down the gauntlet for any view that gives metacognitive abilities a central place.

The philosophical literature on agency and responsibility is replete with such views. Take, for example, Michael Bratman's (1997) influential planning theory. On his view, the ability to construct and enact long-term plans is a central feature of human agency. Inevitably, though, planning of this sort will involve reflecting on mental states that one has currently or that one will need to have in order to implement a plan. If it should turn out that individuals with ASD are impaired in their ability to construct long-term plans, then we are in need of an explanation that

https://doi.org/10.1017/apa.2017.37 Published online by Cambridge University Press

will reconcile their apparent agency with their impairment in the central capacity that Bratman highlights. As a further example, consider Angela Smith's (2005, 2015) 'rational relations' view. On her view, responsibility for various attitudes requires that the attitudes in question be connected to one's evaluative judgments. If this connection requires that one be able to reflectively endorse or renounce the attitudes that one has, then it would seem to require robust metacognitive abilities as well, and it is an open question whether individuals with ASD could be considered responsible agents on this view (in Stout [2017] I argue that the evidence on autism creates a different sort of challenge to Smith's view as well). These are but two examples, but other theorists who have offered accounts of responsibility that require a robust, metacognitive reflective capacity will be vulnerable to the same sort of challenge (e.g., R. Jay Wallace [1994] may also be seen as presenting a metacognitive account).

Moreover, the evidence on metacognition in autism may have implications beyond debates in agency and responsibility. Take, for example, the long-standing debate between moral rationalists and moral sentimentalists. Some rationalists have responded to the sentimentalist claim that moral judgments are grounded in sentiments by appealing to the necessity of our being able to endorse reflectively the attitudes that lead to moral judgments. Christine Korsgaard, for example, writes, 'Our capacity to turn our attention on to our own mental activities is also a capacity to distance ourselves from them, and to call them into question. . . . The reflective mind cannot settle for perception and desire, not just as such. It needs a *reason*. Otherwise, at least as long as it reflects, it cannot commit itself or go forward' (1996: 93). However, if we take this reflective capacity to be necessary for making moral judgments at all, then we should expect autistic persons to be exceedingly bad at making moral judgments, but, as we saw above, that is not at all what is borne out by the evidence.

These concluding remarks are surely more provocative than they are convincing. I do not offer them as a refutation of any of the views mentioned, but as a means of showing that the challenges posed by the evidence from autism research do not stop with the views discussed in this paper. If we are to theorize about human agency and morality, then we need to consider the humanity of so-called marginal agents, and if our theories give us implausible results when doing so, then this is prima facie reason to think that they must be rejected or revised.

NATHAN STOUT
TULANE UNIVERSITY
npstout@gmail.com

References

Baron-Cohen, Simon, Alan M. Leslie, and Uta Frith. (1985) 'Does the Autistic Child Have a "Theory of Mind"?' *Cognition*, 21, 37–46.

Baron-Cohen, Simon, Michelle O'Riordan, Valerie Stone, Rosie Jones, and Kate Plaisted. (1999) 'Recognition of Faux Pas by Normally Developing Children and Children with Asperger

https://doi.org/10.1017/apa.2017.37 Published online by Cambridge University Press

Syndrome or High-Functioning Autism'. *Journal of Autism and Developmental Disorders*, 29, 407–18.
Blair, R. James R. (1996) 'Brief Report: Morality in the Autistic Child'. *Journal of Autism and Developmental Disorders*, 26, 571–79.
Blair, R. James R. (1999) 'Psychophysiological Responsiveness to the Distress of Others in Children with Autism'. *Personality and Individual Differences*, 26, 477–85.
Bratman, Michael. (1997) 'Responsibility and Planning'. *The Journal of Ethics*, 1, 27–43.
Broekhof, Evelien, Lizet Ketelaar, Lex Stockmann, Annette van Zijp, Marieke G. N. Bos, and Carolien Reiffe. (2015) 'The Understanding of Intentions, Desires and Beliefs in Young Children with Autism Spectrum Disorder'. *Journal of Autism and Developmental Disorders*, 45, 2035–45.
Cushman, Fiery. (2013) 'Action, Outcome, and Value: A Dual-System Framework for Morality'. *Personality and Social Psychology Review*, 17, 273–92.
Dziobek, Isabel, Kimberly Rogers, Stefan Fleck, Markus Bahnemann, Hauke R. Heekeren, Oliver T. Wolf, and Antonio Convit. (2008) 'Dissociation of Cognitive and Emotional Empathy in Adults with Asperger Syndrome Using the Multifaceted Empathy Test (MET)'. *Journal of Autism and Developmental Disorders*, 38, 464–73.
Frankfurt, Harry. (1969) 'Alternate Possibilities and Moral Responsibility'. In Frankfurt, *The Importance of What We Care About* (New York: Cambridge University Press, 1998), 1–10.
Frankfurt, Harry. (1971) 'Freedom of the Will and the Concept of a Person'. In Frankfurt, *The Importance of What We Care About* (New York: Cambridge University Press, 1998), 11–25.
Frankfurt, Harry. (1987) 'Identification and Wholeheartedness'. In Frankfurt, *The Importance of What We Care About* (New York: Cambridge University Press, 1998), 159–76.
Frankfurt, Harry. (1992) 'The Faintest Passion.' *Proceedings and Addresses of the American Philosophical Association*, 66, 5–16.
Frith, Uta, John Morton, and Alan Leslie. (1991) 'The Cognitive Basis of a Biological Disorder: Autism'. *Trends in Neurosciences*, 14, 433–38.
Gallagher, Helen and Christopher Frith. (2003) 'Functional Imaging of "Theory of Mind"'. *TRENDS in Cognitive Science*, 7, 77–83.
Hill, Elisabeth L. (2004a) 'Evaluating the Theory of Executive Dysfunction in Autism'. *Developmental Review*, 24, 189–233.
Hill, Elisabeth L. (2004b) 'Executive Dysfunction in Autism'. *TRENDS in Cognitive Sciences*, 8, 26–32.
Hirvelä, Shari, and Klaus Helkama. (2011) 'Empathy, Values, Morality and Asperger's Syndrome'. *Scandinavian Journal of Psychology*, 52, 560–72.
Kennett, Jeanette. (2002) 'Autism, Empathy, and Moral Agency'. *The Philosophical Quarterly*, 52, 340–57.
Korsgaard, Christine M. (1996) *The Sources of Normativity*. Cambridge: Cambridge University Press.
Koster-Hale, Jorie, Rebecca Saxe, James Dungan, and Liane, L. Young. (2013) 'Decoding Moral Judgments from Neural Representations of Intentions'. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 5648–53.
Leslie, Alan M., Ron Mallon, and Jennifer A. DiCorcia. (2006) 'Transgressors, Victims, and Cry Babies: Is Basic Moral Judgment Spared in Autism?' *Social Neuroscience*, 1, 270–83.
Ozonoff, Sally, Bruce F. Pennington, and Sally J. Rogers. (1991) 'Executive Function Deficits in High-Functioning Autistic Individuals: Relationship to Theory of Mind'. *Journal of Child Psychology and Psychiatry*, 32, 1081–105.
Phillips, Wendy, Simon Baron-Cohen, and Michael Rutter. (1995) 'To What Extent can Children with Autism Understand Desire?' *Development and Psychopathology*, 7, 151–69.
Premack, David, and Guy Woodruff. (1978) 'Does the Chimpanzee Have a Theory of Mind?' *Behavioral and Brain Sciences*, 1, 515–26.
Rajendran, Gnanathusharan, and Peter Mitchell. (2007) 'Cognitive Theories of Autism'. *Developmental Review*, 27, 224–60.

https://doi.org/10.1017/apa.2017.37 Published online by Cambridge University Press

Rogers, Kimberly, Isabel Dziobek, Jason Hassenstab, Oliver T. Wolf, and Antonio Convit. (2007) 'Who Cares? Revisiting Empathy in Asperger Syndrome'. *Journal of Autism and Developmental Disorders*, 37, 709–15.

Saxe, R., and N. Kanwisher. (2003) 'People Thinking about Thinking People: The Role of the Temporo-Parietal Junction in "Theory of Mind"'. *NeuroImage*, 19, 1835–42.

Shoemaker, David. (2015a) 'Ecumenical Attributability'. In Randolph Clarke, Michael McKenna, and Angela Smith (eds.), *The Nature of Moral Responsibility: New Essays* (New York: Oxford University Press), 115–40.

Shoemaker, David. (2015b) *Responsibility from the Margins*. New York: Oxford University Press.

Smith, Angela. (2005) 'Responsibility for Attitudes: Activity and Passivity in Mental Life'. *Ethics*, 115, 236–71.

Smith, Angela. (2015) 'Responsibility as Answerability'. *Inquiry*, 58, 99–126.

Stout, Nathan. (2016a) 'Autism, Episodic Memory, and Moral Exemplars'. *Philosophical Psychology*, 29, 858–70.

Stout, Nathan. (2016b) 'Moral Agency and Responsibility: Lessons from Autism Spectrum Disorder'. PhD diss., Tulane University.

Stout, Nathan. (2016c) 'Reasons-Responsiveness and Moral Responsibility: The Case of Autism'. *Journal of Ethics*, 20, 401–18.

Stout, Nathan. (2017) 'Emotional Awareness and Responsible Agency'. *Review of Philosophy and Psychology*. doi:10.1007/s13164-017-0368-x.

Wallace, R. Jay. (1994) *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.

Watson, Gary. (1975) 'Free Agency'. In Watson (ed.), *Free Will*, 2d ed. (New York: Oxford University Press, 2003), 337–51.

Yirmiya, Nurit, Marian D. Sigman, Connie Kasari, and Peter Mundy. (1992) 'Empathy and Cognition in High-Functioning Children with Autism'. *Child Development*, 63, 150–60.

https://doi.org/10.1017/apa.2017.37 Published online by Cambridge University Press