Abstract

The theory of mind (ToM) deficit associated with autism spectrum disorder has been a central topic in the debate about the modularity of the mind. In a series of papers, Philip Gerrans and Valerie Stone argue that positing a ToM module does not best explain the deficits exhibited by individuals with autism (Gerrans 2002; Stone & Gerrans 2006a, 2006b; Gerrans & Stone 2008). In this paper, I first criticize Gerrans and Stone’s (2008) account. Second, I discuss various studies of individuals with autism and argue that they are best explained by positing a higher-level, domain-specific ToM module.

Acknowledgements

I wish to thank Edouard Machery and Arthur Falk for extensive comments on a draft of this paper. Also, I thank Philip Gerrans for his feedback, Shannon Bernard-Adams for many conversations about clinical practice and Autism Spectrum Disorder, and the audience at the 2010 PSA meeting in Montréal.
1. Introduction

In a series of papers, Philip Gerrans and Valerie Stone argue that the theory of mind (ToM) deficit displayed by individuals with autism is not best explained by appealing to a ToM module. Instead, they argue that it is best explained by appealing to low-level deficits, such as gaze processing, that prevent individuals with autism from receiving the same inputs as individuals without autism; thus a domain-general capacity they call “metarepresentation” does not develop as it otherwise would. Their most recent paper (2008) is the most general in scope, and thus I will focus on it.

In this paper, I argue that the Gerrans and Stone are unable to explain key features of the ToM deficit and that the best explanation for it appeals to a domain-specific, higher-level ToM module. First, I discuss Gerrans and Stone’s (2008) explanation, outlining two ways in which it can be interpreted and concluding that neither explains the ToM deficit. Second, I focus on studies of persons with autism’s performance on “reading the mind in the eyes” tests, and I argue that these and studies related to visual processing are best explained by appealing to a domain-specific, higher-level ToM module.

---

1 Gerrans 2002; Stone & Gerrans 2006a, 2006b; Gerrans & Stone 2008.
2. Gerrans and Stone’s Explanation

In this section, I argue that Gerrans and Stone’s (2008) explanation of the ToM deficit can be interpreted in two different ways, which I call the diachronic and the synchronic accounts. They explicitly reference the diachronic account when describing their explanation, but the synchronic account, I argue, is what operates during their explanations of empirical studies. The diachronic account is that the ToM deficit is best explained by lower-level deficiencies present early in development that have a cascading effect on the child with autism throughout development. As Gerrans and Stone argue, “[i]t is their [domain-specific and domain-general systems] interaction in development, rather than the maturation of a mindreading module, which explains the emergence of ToM” (2008, 122). The synchronic account is that related empirical studies are best explained by appealing to current lower-level deficits regardless of subjects’ stage of development.

2.1 The Diachronic Account

Gerrans and Stone (2008, 125) characterize their view as returning to an early explanation of ToM by Baron-Cohen et al. (1985). Baron-Cohen et al. explain ToM by positing a domain-general capacity called metarepresentation that enables ToM as well as other capacities in which a relationship between an object and a representation is represented. Gerrans and Stone’s explanation of ToM is more complex than this early view:

…ToM abilities depend, not on MR [metarepresentation], EF [executive function], language, or recursion per se, but on their developmental interaction
with the low-level precursors previously described, e.g., gaze processing […]

(2008, 128)

Gerrans and Stone are not clear about how appealing to development in addition to “low-level precursors” helps explains ToM. In the following paragraphs, I detail their argument for the diachronic account.

Gerrans and Stone first discuss developmental details with which no one involved in the debate over the explanation of the ToM deficit would disagree. They note that the capacity for ToM depends on the development of many lower-level cognitive mechanisms, such as face processing (2008, 126). The typically-developing child will, in addition to these domain-specific systems, have domain-general systems that develop between the ages of two and four; Gerrans and Stone are interested most in executive function, secondary representation, recursion, and metarepresentation. These “distance us from our primate ancestors […] rather than the fact that we possess some domain-specific ToM mechanism and they do not” (2008, 127).

Next Gerrans and Stone consider how the developmental aspect of their account, of which they provide minimal discussion, might explain persons with autism’s performance on false-belief tests. They argue that the results from such tests are underdetermined since persons with autism’s failure of false-belief tests could result from either lower-level deficits, such as a joint-attention deficit, or from higher-level deficits, such as a deficit in metarepresentation (2008, 129). However, they do think that a mark in their explanation’s favor is that it predicts persons with autism’s success on false-photograph tests and failure on
false-belief tests.\(^2\) Gerrans and Stone argue their account predicts this result because it posits lower-level deficits that affect the individual with autism on the false-belief task but not on the false-photograph task, as the following quotation illustrates:

> Our reply to domain-specific theorists of ToM is that if children with autism have intact MR [metarepresentation], but impaired inputs to MR from impaired low-level domain-specific mechanisms (e.g., joint attention), then one would predict exactly this dissociation between false-photograph and false-belief performance. Individuals with autism perform well on the false-photograph task because the inputs to the metarepresentational system, in that task, are not impaired in autism. (2008, 131)

Note that Gerrans and Stone say nothing about a diachronic developmental interaction in this explanation; instead they appeal only to a lower-level deficit (joint attention).

There is a more general worry for Gerrans and Stone’s diachronic account. Edouard Machery (forthcoming) has recently criticized a similar developmental explanation of Williams Syndrome by Annette Karmiloff-Smith (1998). Machery summarizes Karmiloff-Smith’s related point as follows:

> Brain areas are massively interconnected, so that developmental changes in one of these areas have consequences in numerous other areas. As a result, even minor developmental brain problems have cascading effects, which

\(^2\) Regarding the false-photograph test, see Zaitchik 1990.
affect the development of the whole brain and of all cognitive capacities.

(forthcoming)

Karmiloff-Smith’s argument is similar to the diachronic account that I have sketched. On this account, the ToM deficit is best explained by a lower-level input deficit occurring early in development that results in “abnormal developmental interaction” between that input and higher-level systems (2008, 133).

Machery’s (forthcoming) worry is that though this may seem plausible, it neglects that modularity is a key characteristic of biological development. As Machery notes, “[i]n the context of developmental biology, ‘modularity’ refers to the property that the developmental pathway leading to the development of a trait is shielded from the changes that might happen to other developmental pathways” (forthcoming). Given that we find modularity throughout biological development, it is highly plausible to expect that there is some degree of modularity in the architecture of the mind, even with capacities such as ToM. At this stage, we may infer that Gerrans and Stone’s account is less prima facie plausible than an account that appeals to modularity in higher-level processes.

Gerrans and Stone (2008) do not address this issue, but Gerrans (2002, 310-311) argues that we must determine on a case-by-case basis whether domain-general or domain-specific systems subserve abilities. I grant that we must do this on a case-by-case basis. Nonetheless, considering the ubiquity of modularity in biological development does inform our search for such systems; it directs us to look for modularity. Given the current state of
autism research, the widespread modularity we find in biological development adds plausibility to modular explanations of abilities such as ToM.

As mentioned already, it is unclear that the diachronic aspect of Gerrans and Stone’s (2008) stated view is doing much work in their explanations of empirical findings. Regardless of whether their diachronic account is less plausible than a synchronic account, I will now defend my claim that when Gerrans and Stone (2008) attend to empirical work they do not appeal to “developmental interaction” of any sort; rather, they explain such studies by positing only current lower-level deficits.

If Gerrans and Stone’s (2008) argument hinged on the diachronic component of their account, it would be unclear whether it was better supported than an account appealing only to lower-level deficits. They do not discuss whether the diachronic account explains the ToM deficit better than an account which appeals only to a lower-level deficit such as facial processing. Regardless, when they are discussing empirical studies they do not invoke their diachronic account to explain ToM deficits; thus I suggest that the diachronic account is not doing the explanatory work they claim.

One such example mentioned already is their discussion of persons with autism’s performance on false-belief and false-photograph tests (Gerrans & Stone 2008, 131). In their

---

3 Gerrans and Stone (2008, 128) argue that their account explains the ToM deficit better than one which appeals only to some higher-level deficit, such as executive function, but they do not discuss if appealing only to a lower-level deficit might explain ToM as well.
explanation of these phenomena, they appeal only to “impaired low-level domain-specific mechanisms” and not to any sort of developmental interaction (Gerrans & Stone 2008, 131). As they note: “We believe that deficits on ToM tasks arise essentially from deficits in lower-level input systems” (Ibid.). Gerrans and Stone (2008, 132) claim that appealing to low-level deficits explains persons with autism’s performance more parsimoniously, but they provide no discussion of how such parsimony relates to the diachronic aspect of their account.

Gerrans and Stone (2008, 132-133) also explain the performance of higher-functioning individuals with autism by appealing only to lower-level deficits. This explanation is discussed at length below, but it is worth noting that after discussing persons with autism’s performance on false-belief and false-photograph tests and after discussing the performance of higher-functioning persons with autism on other tasks, Gerrans and Stone claim that their developmental interaction account provides the best explanation for both phenomena, as the following quotation illustrates:

The weight of evidence indicates that the ToM impairments in autism… are a result of abnormal developmental interaction between low-level systems and higher-level domain-general capacities, caused essentially by deficits in ToM precursors. (2008, 133)

Since the explanations they provide of both of the aforementioned phenomena rely only on their appeal to lower-level deficits, it is unclear why developmental interaction should play a role in explaining ToM deficits. Since Gerrans and Stone (2008) do not mention it at all in
their explanations of these two crucial phenomena, it does not seem that the diachronic component of their account is what does the real explanatory work.

Even if it weren’t the case that the work in Gerrans and Stone’s (2008) argument is being done by appealing only to current lower-level deficits, it seems that such an explanation of the ToM deficit still threatens the explanation that appeals to a domain-specific ToM module. That is, aside from being interested in interpreting Gerrans and Stone charitably, it is important for my argument that I deal with the possibility of explaining ToM deficits by appealing only to current lower-level deficits.

2.2 The Synchronic Account

Even if the likelihood of modularity in a higher-level process like ToM is increased by recognizing the near ubiquity of modularity in biological development, Gerrans and Stone’s explanation of key studies in cognitive neuroscience still threatens a modular explanation of ToM deficits. Without relying on developmental interaction, Gerrans and Stone argue that studies of higher-functioning persons with autism where subjects demonstrate difficulty on ToM tasks can be explained by appealing only to lower-level deficits. In this section, I examine several such studies, including recent studies not discussed by Gerrans and Stone.
Gerrans and Stone (2008, 132-133) discuss three case studies of individuals on the autism spectrum who were very high-functioning; two were college students (a computer science student and a physics student), and the other was a mathematics professor who had won the Fields Medal (Baron-Cohen et al. 1999). Gerrans and Stone discuss this study because it is a potential counterexample to their explanation of ToM deficits. This study is a potential counterexample because, given their careers, these subjects have an advanced ability “to represent relations between symbols and their objects (magnitudes, abstract spatial relations) and to perform recursive computations over these symbols” (Gerrans & Stone).

4 I will not discuss false-belief tests further since it is unclear whether they are best explained by a domain-specific ToM module or not given that many individuals with autism pass one kind of false-belief test, the first-order belief test, but fail another, the second-order belief test (Stone et al. 1998). Furthermore, there is a large literature disputing whether these tests are measuring ToM at all (e.g., Egeth & Kurzban 2009; Bloom & German 2000).

5 Baron-Cohen et al. (1999, 479) note that these three individuals were diagnosed with Asperger syndrome but recognize that many of its symptoms overlap with those of higher-functioning autism; regardless of controversy over whether Asperger syndrome is distinct from higher-functioning autism (Klin et al. 1995, 1129), these individuals are on the autism spectrum and, most importantly for the present discussion, they have ToM deficits.
Since on Gerrans and Stone’s account ToM is supported by domain-general processes, individuals who are able to engage in activities requiring metarepresentation, recursion, and executive function, such as the activities in which these three subjects are able to excel, while being unable to perform well at activities requiring the representation of others’ mental states would provide a counterexample to their explanation.

Baron-Cohen et al. (1999) administered three tests: a folk psychology test (ToM), a folk physics test (basic problem solving about the physical world), and an executive function test (Tower of Hanoi test). The first test is the most relevant to the present discussion. Subjects were shown 36 photographs of the eye region and asked to choose the word that best described what the person in the photograph was thinking or feeling (Baron-Cohen et al. 1999, 479); this is called the “reading the mind in the eyes” test (for discussion, see Baron-Cohen et al. 1997; Baron-Cohen et al. 2001). The subjects performed well on the executive function and folk physics tests, but all exhibited deficits on the ToM test.

Gerrans and Stone (2008, 133) argue that since the ToM test in Baron-Cohen et al.’s (1999) study isolated the eye region, it is more likely that the subjects’ difficulty in identifying emotions was due to a lower-level deficit in facial or gaze processing:

[...] all three had difficulty in inferring what someone was feeling, or paying attention to, from pictures of the eye region of the face (Baron–Cohen et al. [1999]), indicating a problem with lower-level domain-specific capacities for face and gaze processing rather than MR [metarepresentation]. (2008, 133)
There is a problem with their interpretation of this study, and there are also related studies that weigh against Gerrans and Stone’s interpretation. Note that, as with their explanation of performance on the false-belief tests mentioned above, Gerrans and Stone do not discuss how their diachronic account contributes anything to their explanation of the behavior of these subjects.

The problem with Gerrans and Stone’s (2008) explanation is that they neglect to mention the control task administered to the subjects. To rule out a facial or gaze processing deficit, the subjects and the control group were asked to identify individuals’ gender by viewing only their eye region. On this task, the three subjects performed at the same level as the control group, but all three scored greater than one standard deviation below the mean of the control group on the ToM task (Baron-Cohen et al. 1999, 479-480). Since the subjects performed as well as the control group on a general facial processing task not related to emotion, the ToM deficits exhibited by these subjects are not likely due to a lower-level deficit related to facial or gaze processing.

Surprisingly, Gerrans and Stone (2008) do not address this aspect of the study, but one can imagine how their response might go. They might argue that the control task of identifying gender from a photograph of someone’s eye region does not decisively rule out that ToM is supported by a higher-level, domain-general process such as metapresentation. Furthermore, they might grant that the control task does rule out that the subjects were generally impaired at facial processing, but it does not rule out that their poor performance on the ToM task could have been due to a more specific, lower-level impairment affecting only
the facial processing of emotions. That is, though some elements of facial processing may be intact (i.e., processing of facial features sufficient enough to determine someone’s gender), there is a domain-specific, lower-level component of facial processing that provides inputs of facial emotions.

This potential reply from Gerrans and Stone argues that the control task does not allow us to determine whether the ToM deficits in the Baron-Cohen et al. (1999) study are due to the failure of a higher-level, domain-specific module (a ToM module) or of a lower-level, domain-specific module (an emotion-only facial processing module) that is failing to provide inputs to a domain-general process (e.g., metarepresentation). Recent studies in facial processing that I will now discuss lend support to the former explanation over the latter.

Two recent studies (Lahaie et al. 2006; Humphreys et al. 2007) of facial processing by higher-functioning individuals on the autism spectrum suggests that, though some lower-functioning persons with autism have difficulty with facial processing, in higher-functioning individuals ToM deficits may be dissociated from facial processing deficits. Lahaie et al.’s (2006) study examined a group of higher-functioning persons with autism using two tests to determine whether their configural facial processing differed significantly from the control group and whether they responded to the priming of facial parts like the control group did. For the first test, Lahaie et al. used a standard design whereby faces were inverted and participants were asked to match upright and inverted faces. Face inversion is known to
disrupt recognition of faces because it inhibits normal configural processing of facial features.

In previous studies of the face inversion effect, persons with autism performed better than controls in recognizing inverted faces but similarly or worse than controls with upright faces, causing some researchers to argue that persons with autism have a deficit in configural facial processing (Lahaie 2006, 31). In Lahaie et al.’s 2006 study, the group of subjects with autism was shown both faces with neutral expressions (to prevent a potential confound from faces showing emotion) and non-facial stimuli, Greebles, that would also be inverted. Given previous studies, Lahaie et al. expected that the group of subjects with autism would not show an inversion effect, which would support the deficiency in configural face processing posited by previous studies (Lahaie et al. 2006, 31). Instead, both the control group and the group of subjects with autism exhibited the face inversion effect, which supports the view that higher-functioning persons with autism do not have a deficit in configural facial processing; this finding is consistent with a similar study by Joseph and Tanaka (2003).6

The second test administered to the subjects in Lahaie et al.’s (2006) experiment was designed to determine whether the group of subjects with autism would exhibit a priming effect similar to the control group. I will not discuss this second test in detail; the relevant result is that the group of subjects with autism performed similarly to the control group on

6 For discussion of the discrepancy between their study and previous findings, see Lahaie et al. (2001, 33).
the priming task, suggesting that individuals in both groups benefited from configural information being made available (Lahaie et al. 2006, 38). These findings support the view that configural facial processing is not impaired in individuals on the autism spectrum, but one might ask how configural facial processing is relevant to the “reading the mind in the eyes” test. It is relevant because Lahaie et al. (2006, 38) found that, contrary to previous studies, subjects with autism looked at the eye region of the face more than other areas of the face. Not only did these subjects pay more attention to the eye region, but moreover they processed the eye region better after being primed with that area than any of the other facial areas with which they were primed.

Their overall performance at equal with the control group and their enhanced ability to process the eye region suggest not only that these higher-functioning subjects with autism are not suffering from a lower-level facial processing deficit but also that they process the eye region better than other facial regions. This lends validity to the methodology in the “reading the mind in the eyes” test and also weighs against Gerrans and Stone’s appeal to lower-level facial processing deficits to explain the ToM deficits in Baron-Cohen et al. (1999). Recent experiments on auditory processing by persons with autism also lend support

7 Humphreys et al.’s (2007) study had similar findings. To rule out a lower-level deficit, subjects compared facial expressions. There was no significant difference between the two groups of subjects in the ability to make fine-grained facial comparisons, suggesting there was no lower-level facial processing deficit.
to the view that difficulty identifying emotions in spoken voices is not due to lower-level deficits.  

Gerrans and Stone could respond to these studies in two different ways. First, they could argue that though the studies discussed above seem to indicate that higher-functioning individuals can have ToM deficits in the absence of lower-level deficits, this does not rule out the possibility of them having a higher-level deficit, such as with executive function. Second, they could accept the findings of these experiments but then posit a specialized lower-level module devoted only to processing emotional information in facial processing (and another such module for auditory processing; see fn. 8). Neither of these responses will work.

The first reply will not explain the behavior of the three subjects in Baron-Cohen et al.’s (1999) study. The subjects were tested not only for ToM using the “reading the mind in the eyes” test but also for executive function capabilities using the Tower of Hanoi test. Gerrans and Stone could not appeal to the domain-general process of metarepresentation that they posit, since these three individuals are clearly skilled at advanced representation and at

---

8 Subjects with autism in Rutherford et al.’s (2002, 190) auditory processing study were significantly impaired in identifying the mental states of speakers from recordings but were not in assigning speakers to an age range (control task). Kleinmann et al. (2001) had similar findings (the control task was identifying a speaker’s gender). See also Golan et al. (2006) and Golan et al. (2008); the former uses videos rather than static images.
recursive computations, given their professions, as Gerrans and Stone admit (2008, 133). Furthermore, though it makes use of false-belief tests, a study of a subject with damage to the left amygdala suggests that ToM is dissociated from executive function (Fine et al. 2001), and there are numerous other studies suggesting that young children diagnosed with autism do not exhibit deficits in executive function (e.g., Griffith et al. 1999; Yerys et al. 2007). As a result, it seems that Gerrans and Stone cannot appeal to a higher-level, domain-general process to explain the subjects’ performance in Baron-Cohen et al.’s (1999) study.

The second reply may seem plausible but if Gerrans and Stone went this route they would strongly detract from the force of their argument. Gerrans and Stone (2008) emphasize that they are seeking a more parsimonious account of the ToM deficit by appealing to lower-level deficits. Their desire for parsimony is clear from the cautionary phrase that they co-opt: *modulae non sunt multiplicanda praeter necessitate* (2008, 122). To explain the behavior of subjects with autism in the visual and auditory processing tasks (see fn. 8), Gerrans and Stone would need to posit a deficit in the module for emotion processing for vision and for hearing (while all other lower-level processing modules are intact). Rather than explaining these deficits by adding a module for each modality, these studies can be explained by appealing to a single higher-level, domain-specific ToM module that accounts for deficits in both visual and auditory tasks.

---

9 For a review of the large literature on the relationship between executive function deficits and autism, see Hill 2004.
3. Conclusion

In this paper, I argued that a modular explanation of the ToM deficit in persons with autism is more plausible than the explanation provided by Gerrans and Stone (2008). The plausibility of the modular account is supported by its ability to provide an explanation of studies of higher-functioning individuals with autism and by the fact that modularity is expected given its ubiquity in biological development.
References


