"Explaining the Theory of Mind Deficit in Autism Spectrum Disorder," *Philosophical Studies*, forthcoming.

Marcus P. Adams
University of Pittsburgh
Department of History & Philosophy of Science
1017 Cathedral of Learning
Pittsburgh, PA 15260 USA
Email: mpa9@pitt.edu

Abstract

The theory of mind (ToM) deficit associated with autism has been a central topic in the debate about the modularity of the mind. Most involved in the debate about the explanation of the ToM deficit have failed to notice that autism's status as a spectrum disorder has implications about which explanation is more plausible. In this paper, I argue that the shift from viewing autism as a unified syndrome to a spectrum disorder increases the plausibility of the explanation of the ToM deficit that appeals to a domain-specific, higher-level ToM module. First, I discuss what it means to consider autism as a spectrum rather than as a unified disorder. Second, I argue for the plausibility of the modular explanation on the basis that autism is better considered as a spectrum disorder. Third, I respond to a potential challenge to my account from Philip Gerrans and Valerie Stone's recent work (Gerrans 2002; Stone & Gerrans 2006a, 2006b; Gerrans & Stone 2008).

Keywords

modularity, theory of mind, autism spectrum disorder, spectrum disorders, psychological theory

"Explaining the Theory of Mind Deficit in Autism Spectrum Disorder"

0. Introduction

Autism has been a central topic in the debate about the modularity of the mind (e.g., Carruthers 2006a, 18; Samuels 1998, 597). Most involved in the debate do not question whether there are modules for some lower-level input systems like vision (e.g., Fodor 1983). What is at stake is how *much* of the mind is modular. For autism, the question is whether the theory of mind (hereafter ToM) deficit displayed by persons with autism is best explained by appeal to some domain-general process or to a domain-specific ToM module. In this paper, I argue that the shift from viewing autism as a unified syndrome to a spectrum disorder increases the plausibility of the explanation of the ToM deficit that appeals to a domain-specific, higher-level ToM module. First, I discuss autism considered as a spectrum disorder rather than as a unified syndrome and then discuss briefly how I will be using 'module'. Second, I argue for the plausibility of the modular explanation on the basis that autism is better considered as a spectrum disorder. Third, I respond to a potential challenge to my account from Philip Gerrans and Valerie Stone (Gerrans 2002; Stone & Gerrans 2006a, 2006b; Gerrans & Stone 2008), focusing on their most recent work (2008).

1. Autism and Modularity

1.1 Autism as a Spectrum

The *DSM-IV-TR* (hereafter *DSM*) defines autism as a pervasive developmental disorder (PDD) characterized by impairments in social interaction and communication as well as by "restricted repetitive and stereotyped patterns of behavior, interests, and activities" (APA 2000, 75). Other

PDDs include Rhett's Disorder, Childhood Disintegrative Disorder, Asperger's Disorder, and

PDD not otherwise specified (PDD-NOS). The first two impairments that characterize autism,

i.e., social interaction and communication, are qualitative impairments, i.e., impairments

differing in kind from abilities exhibited by individuals who have undergone typical

development.

To be diagnosed with autism an individual must exhibit a total of six behavior types from

the twelve listed in the *DSM*, including at least two from the first group (social interaction) and at

least one from the second (communication) and one from the third (restricted repetitive and

stereotyped patterns). The ToM deficit is not presently an explicit component of the diagnostic

criteria, but these criteria do relate to this deficit (e.g., difficulty in pretense).[1]

Though the *DSM* identifies autism as a disorder, it is important to urge caution to those

who act as if it were a unified syndrome with a high degree of similarity among individuals

diagnosed with it. The potential for a great degree of variance among those diagnosed with

autism should be evident from merely examining the possible combinations of the twelve

behavior types that individuals with autism can have, but this variance has not always been

recognized.

Early work by Leo Kanner (1943) treated autism as a unified syndrome. Kanner posited

that the children he studied came "into the world with innate inability to form the usual,

biologically-provided affective contact with people…" (Kanner 1943, 250). Since Kanner's early

work, researchers have recognized the diversity among individuals with autism. Autism is now

better described as a spectrum disorder rather than a unified syndrome.[2] But a spectrum disorder

---

[1] For additional diagnostic criteria, see the *DSM-IV* (APA 2000, 75).
[2] Psychiatrists distinguish between categorical and dimensional classificatory systems. The
DSM-IV is a categorical system that "works best when all the members of a diagnostic class are

is not simply a continuum from more to less severe cases of a common deficit or set of deficits; rather, autism is a complex disorder better conceptualized in terms of combinations of various symptoms than as a unified syndrome (Wing 1988). This shift to focus on symptoms is not surprising, since similar transitions have occurred in the history of psychiatry with other disorders, e.g., schizophrenia (Coltheart & Langdon 1998).

Viewing autism as a spectrum rather than as a unified syndrome has important methodological implications for psychological theory and clinical practice as well as for the plausibility of the explanation of the ToM deficit that I will advance, and I will discuss these below. For now, it is important to highlight that given autism's characterization as a spectrum it should be possible, in principle, to find any one of the three main areas of deficits, as well as the individual sub-areas, dissociated from the others (see Bishop 2000, 259-260). Such higher-functioning individuals would have deficits in only one or two of the three major deficit areas listed in the *DSM*, displaying some of the behaviors of those on the autism spectrum but not falling under the autism spectrum category. Such individuals are not, in fact, uncommon (Bishop 2000, 260).

Asperger's disorder is one such case of individuals who share some but not all symptoms with individuals diagnosed with autism. Many individuals diagnosed with PDD-NOS do as well. The purpose of this paper is not to enter into the controversy surrounding differentiating Asperger's disorder from higher-functioning autism (Klin et al. 1995; Klin & Volkmar 2003). This controversy concerns whether the *DSM* should give Asperger's disorder its own category.[3]

---

homogenous, when there are clear boundaries between classes, and when the different classes are mutually exclusive" (APA 2000, xxxi). Dimensional systems, however, are for disorders on continua. The current *DSM-V* task force is investigating how the "concept of spectrum disorders" can work alongside the traditional Axis I-Axis II distinction (Kupfer et al. 2002, xx).

[3] See Simon Baron-Cohen's (2009) recent *New York Times* commentary.

However, most in this debate agree that Asperger's is "on a phenomenological continuum with autism" (Klin et al. 1995, 1129). That is, most agree that the difference between autism and Asperger's is a difference in *degree*.[4] The dispute concerns whether the differences in degree between higher-functioning persons with autism and those with Asperger's disorder are sufficient to warrant placing the latter within its own diagnostic category. My point is that whether one uses the label 'Asperger's' or 'autism', there are individuals—those diagnosed with Asperger's—who share the majority of the observed behavioral deficits exhibited by individuals with autism, since the two disorders are currently differentiated only by level of intelligence and the onset of language. My focus is on subjects among these individuals who exhibit a ToM deficit in the absence of many other deficits.

1.2 'Module' and Higher Cognitive Processes

The term 'module' is used in various ways in the literature, so it will be useful briefly to explain how I use it. Fodor's influential book *The Modularity of Mind* (1983) argued that lower-level input systems, e.g., vision, are modular but that "central systems" are not, e.g., those that fix perceptual belief. Fodor included only these peripheral input systems in his account of modularity because of the manner in which he characterized 'module'. He identified nine characteristics of modular systems; key among these was the notion of "informational encapsulation," which Fodor took to be the most important feature of modularity (1983, 37). A module is informationally encapsulated "if it operates on proprietary input and ignores available information that is relevant to the computation at hand" (Carey 2009, 9). The well-known

---

[4] Baron-Cohen (2009), however, argues that this is still an open question because there has been insufficient research into possible biological differences between individuals with autism and with Asperger's.

Müller-Lyer illusion is a common illustration of such encapsulation. Much of the mind is not modular on Fodor's (1983) account.

More recently, evolutionary psychologists have proposed what many call the massive modularity thesis (MMT) (Cosmides & Tooby 1992; Tooby & Cosmides 1992; Machery & Barrett 2006). On the MMT, many of the abilities that Fodor (1983) argued were subserved by domain-general systems are subserved by domain-specific modules. On one such representative view,

> […] humans have a faculty of social cognition, consisting of a rich collection of dedicated, functionally specialized, interrelated modules […] organized to collectively guide thought and behavior with respect to the evolutionarily recurrent adaptive problems posed by the social world. (Cosmides & Tooby 1992, 163)

One important difference between Fodor's (1983) account and the above interpretation of the MMT, beyond that the latter argues that the mind is much more modular, is that for the latter whether a module is informationally-encapsulated depends on its particular functional specialization (Barrett & Kurzban 2006, 629).[5]

I must make two further distinctions regarding the MMT: first, it is independent from claims of innateness; and second, a module individuated in terms of its functional specificity does not imply that there is a single, distinct anatomical realization of that capacity in the brain.

---

[5] Critics of MMT often presuppose without argument what they take to be Fodor's view about informational encapsulation, i.e., that it is a necessary condition, leading them to argue that domain specificity would most often not be adaptive (see, e.g., Gerrans 2002, 310). Many have taken Fodor (1983) to be arguing that informational encapsulation is a necessary condition for a module, but Max Coltheart (1999) argues that on Fodor's account informational encapsulation was only a characteristic or "typical" feature of a modular system and, in fact, Fodor does argue that the "notion of modularity ought to admit of degrees" (Fodor 1983, 37).

Some favoring the modular explanation for the ToM deficit in persons with autism have supposed that this module is innate (cf. Baron-Cohen 1998, 182-184). But the version of the MMT assumed in this paper is compatible both with the view that argues that such modules are, at least in part, innate (Baron-Cohen 1998, 184; Leslie 1987, 424) and with the view that modules arise through a complex developmental path (see Machery forthcoming; *contra* Karmiloff-Smith 1998).[6] The MMT is designed to explain the cognitive architecture of the mind regardless of whether it arises via innate modules or developmentally-constructed modules (see Barrett & Kurzban 2006, 638).

Second, the view of 'module' that I assume does not require that each functional module be realized in a single area of the brain. Instead, a functional module could be distributed across areas in the brain that together contribute to that function. Some make inferences from functional modularity to anatomical modularity (see Bergeron 2007; Carrington & Bailey 2009), but this additional step is not required to argue that the ToM deficit in autism is best explained by a higher-level, domain-specific ToM module.

2. Theoretical and Clinical Implications of Considering Autism as a Spectrum

I argue that viewing autism as a spectrum disorder instead of as a unified syndrome increases the plausibility of the modular explanation of the ToM deficit. Many philosophical discussions of autism fail to recognize the significance of conceiving of autism as a spectrum, so the present section is designed to fill this lacuna.

---

[6] However, Karmiloff-Smith (2009, 59) has recently argued that her "neuroconstructivist" account accommodates modularity in the adult brain, insisting that it would arise during the "ontogenetic process of gradual modularization."

2.1 Diagnosing, Intervening, and Discovering

I will now distinguish three different goals for which one could use studies of individuals with autism: developing diagnostic criteria, designing clinical interventions, and discovering cognitive architecture. For each I will highlight the inference involved, but the first goal and the last goal are most relevant to my argument in the next subsection. In discussing these three goals, I will focus on two types of studies in which one could engage. The first type examines an individual or individuals with a specific deficit, and the second examines a heterogeneous group with overlap in key deficits.

If one wants to develop valid criteria for diagnosing individuals with autism, the latter type of study will help provide useful diagnostic guidelines for clinicians. Studying such a group with overlapping symptoms will help you discover what characteristics, on average, will result in successful diagnoses in clinical settings, as I will discuss in examples below. The goal in developing and evaluating such criteria is to provide clinicians with broad categories that will capture a range of behaviors. A successful diagnosis by means of such categories occurs when a patient is identified as exhibiting behaviors fitting within the range of behaviors of a certain disorder X and, after being diagnosed, the disorder X is predictive of that patient's future behaviors. We must use group studies to develop such criteria because we want to provide clinicians with the average range of behaviors and symptoms in the population so that they can identify these in patients.

One widely accepted way of evaluating the validity of diagnostic criteria is to conduct large-scale, longitudinal studies of individuals and determine if the criteria reliably predict future episodes of the behavior in question as well as its occurrence in co-twins (Robins & Guze 1970). Attempts to clarify the *DSM* diagnostic criteria for major depressive disorder have done just this.

For example, Kendler & Gardner (1998) studied 2,163 female subjects to determine whether several of the current major depressive disorder criteria, i.e., number of symptoms, duration, and level of severity or impairment (see APA 2000, 369ff), would accurately predict future episodes of major depressive disorder as well as the risk for the disorder in a co-twin.

Kendler and Gardner found that criteria different from or below the *DSM-IV* threshold criteria were more useful than those in the *DSM-IV*: "Syndromes that met fewer than five criterion A symptoms, lasted for less than 2 weeks, or were formed of symptoms that were quite mild or produced no impairment had considerable predictive and familial validity" (1998, 174). When arguing that a set of diagnostic criteria is valid, like in this example, one makes an inference from their predictive power in group studies to their predictive power in the general population (an inference that moves from a claim about a group to a claim about a group). In the next subsection, I will discuss how researchers have criticized and refined the diagnostic criteria for autism in much the same way.

When intervening in a clinical setting it is necessary to catalogue each individual's symptoms when he or she presents for treatment, but one still needs broad categories in guidebooks like the *DSM* for diagnosing. Klin et al. make a similar point relating to Asperger's:

> Given the prevailing difficulties in the definition of [Asperger's] and the great heterogeneity of the condition, it is crucial that the aim of clinical assessment be a comprehensive and detailed profile of the individual's assets, deficits, and challenges, rather than simply a diagnostic label. (2003, 4)

These two foci—cataloguing individuals' symptoms for interventions and using broad diagnostic categories based upon large-scale, longitudinal studies—are obviously not mutually exclusive. My point is to emphasize that one must engage in the first when *intervening* but in the second

when developing criteria for *diagnosing*. When one intervenes one makes an inference from a particular patient's array of deficits and life situations, and the interplay between these, to what interventions will be successful for that patient (an inference that moves from a claim about an individual to a claim about that same individual).

One might argue that in designing an intervention one really seeks a maximally specific class, given what one knows. I have characterized this sort of inference as an individual-to-same-individual inference because there are often many historically contingent factors relevant to the intervention a clinician crafts (e.g., the loss of one's job during an unanticipated economic recession right after one's child contracts a terminal illness). In the case of such unique historical contingencies, the maximally specific class would have only one member, i.e., the particular patient. If we catalogued only a patient's *psychological* deficits, viewing the crafting of interventions as seeking a maximally specific class would seem a correct characterization since there would be other members in such a class who would require the same intervention (as with Hempel's example of penicillin-resistant streptococcus; see Salmon 2006, 54ff). However, since we take into account not only the patient's psychological deficits but also the patient's social and economic situation, many interventions must be tailored specifically for one patient and will not generalize to other patients.

To discover if the mechanism subserving a capacity such as ToM is domain-specific modules (and dissociable from other capacities) one should engage in a study of individuals with a specific deficit that they have in (near) isolation. Since autism is not a unified syndrome, if we studied mixed groups and looked for main effects from averages of their behavioral performance, we would lose information relevant to discovering whether capacities such as ToM are dissociated from other higher-level capacities or lower-level deficits. There is a large literature

devoted to the inferences made from single-patient studies (e.g., Caramazza 1986; Caramazza & McCloskey 1988; Glymour 2001, ch. 11). I am not advancing the view held by Caramazza and colleagues where *only* single-patient studies are appropriate for making inferences to cognitive architecture, i.e., what some call the "radical view" (Robertson et al. 1993). To discover underlying cognitive architecture, one should make inferences from studies of individuals or groups of individuals who share unique profiles where one deficit is present in the absence of others. Doing this requires that one have rigorous screening protocols to be sure there is a high degree of overlap among subjects (Robertson et al. 1993, 716).

We may find only a few individuals who exhibit a ToM deficit in the absence of nearly all of the other deficits associated with autism, but this is not problematic for making an inference from these cases to a claim about cognitive architecture. Finding just one case would license such an inference. The form of this inference is as follows:

1. Individual A has a deficit of capacity X and does not display any other (salient) deficits;

2. Given the absence of other deficits, capacity X in individual A must be dissociated from other capacities and subserved by a domain-specific module;

3. Thus, capacity X must, in individuals B, C, D, […], be dissociated from other capacities and subserved by a domain-specific module.

This is an inference that moves from a claim about an individual, or possibly a group with a high degree of similarity in their capacities and deficits to a general claim about human cognitive architecture. Whether deficit X is, in fact, often or even *almost* always comorbid with deficits Y and Z is irrelevant for judging whether deficit X is caused by the failure of a dissociable piece (module) of cognitive architecture. If we attempted to make such an inference from a group of

heterogeneous individuals, e.g., a group where all members had a deficit of capacity X but many also had comorbid deficits in capacities Y, Z, P, or Q in various combinations, we would be unable to determine whether X is subserved by a dissociable module or by some combination of the other capacities of which the subjects also exhibited deficits.

Researchers in cognitive neuroscience make a similar type of inference when examining individuals with a nearly isolated deficit to expose which brain structures subserve particular capacities. For example, research with amnesic patients who have damage to the hippocampal system suggests that declarative memory is dissociated from other types of memory, including memory related to skilled performances (for discussion, see Adams 2009). Although these cases often involve identifying behavioral deficits associated with damage to a particular anatomical area such as the hippocampal system, the point of analogy to autism remains—an inference from a case of an individual's isolated deficit allows us to infer what the underlying structure is for others without that deficit. The famous case of H.M. is an example of such an inference (Scoville & Milner 1957).

2.2 Explaining the ToM Deficit and Autism's Status as a Spectrum Disorder

I now want to connect these methodological issues to my argument for the modular explanation of the ToM deficit. Conceiving of autism as a spectrum makes the modular explanation of ToM deficits in individuals with autism more plausible than a non-modular explanation. Strangely, most philosophers and psychologists have neglected to consider that viewing autism as a spectrum has implications for the level of plausibility we should assign to the modular explanation of the ToM deficit. I sketch below the plausibility argument that I will provide in this section:

Plausibility Argument from the Claim that Autism is a Spectrum Disorder

      1. If we had cases of individuals with a ToM deficit in *complete* isolation, this would settle the debate in favor of the domain-specific, higher-level ToM explanation;

      2. We do not have a study of any individual fitting the description in #1;

      3. However, we do have sufficient empirical support for identifying autism as a spectrum disorder, and each symptom of a spectrum disorder can occur independently of the other symptoms;

      4. From #3 it follows that it is plausible that in some individuals the ToM deficit occurs in isolation, though perhaps infrequently;

      5. From #4 it follows that the most plausible explanation for ToM capacity appeals to a domain-specific, higher-level ToM module.

Researchers have come close to finding higher-functioning individuals who possess the ToM deficit in near isolation, but no one to my knowledge has studied an individual with this deficit in *complete* isolation of other salient deficits. The three individuals in Baron-Cohen et al.'s (1999) study, which I will discuss in more detail below in the next section, were recognized as a "rare opportunity" due to the near isolation of their ToM deficit, but even these individuals had deficits in other areas related to the autism spectrum. For example, Baron-Cohen et al. (1999) report that subject GC had some level of social interaction deficit. GC had difficulty understanding social norms, and in college he would physically remove anyone using a computer when he desired to use it (Baron-Cohen et al. 1999, 476-477). Thus, GC and the other two subjects in this study only approximate subjects with an isolated ToM deficit.

Even in the absence of studying such individuals, we can evaluate the plausibility of the modular explanation of ToM by looking at autism's status as a spectrum disorder. Large-scale empirical studies guided the transition from viewing autism as a uniform syndrome to a spectrum. The current guidelines for diagnosing autism in the *DSM-IV* are a significant improvement over those in the *DSM-III-R*, the latter of which had diagnostic criteria that were

too broad and resulted in a large number of false positives (Factor et al. 1989; Volkmar et al. 1992). Volkmar et al. (1994) tested possible definitions for the *DSM-IV* on 977 subjects, helping confirm that the *DSM-IV* criteria had much more predictive power and sensitivity than previous criteria.

Volkmar et al.'s (1994) field test established that the *DSM-IV* criteria are highly successful at diagnosing individuals with "classic autism." However, as Tanguay et al. note, there are many "individuals who present with symptoms clearly resembling those of the more classic autistic cases, but for whom the diagnosis cannot be made" (1998, 272). These individuals fall outside the current criteria because their symptoms are either not severe enough or because their symptoms fall into only one or two of the three domains required for an autism diagnosis. Numerous large-scale studies have led many to hold that autism is best understood as a spectrum and should be categorized dimensionally, not categorically (for this distinction, see fn. 3; for recent treatment with a study of 140 children and adolescents, see Kamp-Becker et al. 2010). The inference made in these studies to identify autism as a spectrum is an instance of the kind discussed in the subsection above under diagnosing.

Most individuals on the autism spectrum will be grouped together around commonly-held sets of deficits, but individuals with "subthreshold symptomatology," who exhibit deficits in only two out of the three symptom domains, are not at all uncommon (Bishop 2000, 259). There also will be outliers who have symptoms from only one of the three domains. Since autism is a spectrum disorder, any of its symptoms can occur in isolation of each other, as psychologist Dorothy Bishop argues:

"[…] there is no necessary association between the three domains of impairment that characterize autism: They can be dissociated, especially in higher-functioning individuals. (Bishop 2000, 259)

Psychiatrist Lorna Wing makes a similar point: "[…] psychological impairments in autism and related conditions can vary independently of each other. Also, any combination may be found…" (Wing 1988, 98). Thus, it is likely that there will be outliers that exhibit only a ToM deficit and no others, though such cases may occur very infrequently.

We are now at a point where we can evaluate whether the explanation that appeals to a domain-specific, higher-level ToM module is more plausible than other potential explanations. I argue that we may legitimately infer that the ToM modular explanation is more plausible than other explanations from the likelihood that there are individuals with a ToM deficit and no other related deficits. If we located such an individual, then studying him or her would decide the matter, but since we have not done so we must make the inference from the likelihood that such an individual exists. The form of this inference is similar to the kind described under discovering cognitive architecture in the previous subsection, but we are able to infer only the *plausible* underlying architecture from the *likely* existence of the individual with an isolated ToM deficit. Thus, I argue that the domain-specific, higher-level ToM module explanation is more plausible than a non-modular explanation, such as the one I will discuss in the next section.

3. Can We Explain the ToM deficit by Appealing only to Lower-Level Deficits?

In a series of recent papers, Philip Gerrans and Valerie Stone (hereafter G&S; Gerrans 2002; Stone & Gerrans 2006a, 2006b; G&S 2008) argue for a non-modular explanation of ToM. Such

an account threatens the explanation that I offer in the previous section. In this section, I provide

reasons to reject their account, which they describe in the most detail in their 2008 paper.[7] This

section is not designed merely to criticize G&S (2008) in the interest of shoring up my

explanation; rather, throughout this section I provide positive empirical evidence for the modular

explanation of ToM by appealing to recent studies on auditory and visual processing that G&S

(2008) do not discuss.

G&S (2008) hope to explain the ToM deficit by appealing to lower-level deficiencies

present early in development that have cascading effects throughout the development of the child

with autism. G&S make this claim in the following quotation: "[i]t is their [i.e., 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; emphasis added).

Although they claim that development is a key component of their explanation, Gerrans and

Stone do not, in fact, appeal to *development* when they attempt to explain several empirical

studies (I outline this interpretational claim in more detail in Adams forthcoming). Instead, when

they attend to the results of empirical studies they appeal only to *lower-level deficits* when

explaining subjects' ToM behavior.

Given that development does not play a role in their explanations of relevant empirical

studies, I propose to interpret G&S as offering what I call a synchronic explanation of the ToM

deficit.[8] On the synchronic account, ToM deficits are best explained by appealing *only* to lower-

level deficits in persons with autism regardless of their stage of development (e.g., facial

---

[7] I outline my criticisms of G&S's (2008) account in more detail in Adams (forthcoming).
[8] In Adams (forthcoming), I provide additional worries facing a diachronic developmental
account such as what Gerrans and Stone (2008) view themselves as providing. For similar
concerns related to Karmiloff-Smith's (1998) developmental explanation of Williams Syndrome,
see Machery (Forthcoming).

processing deficits). I will now examine a key study that G&S discuss and several that G&S do not discuss, arguing that they are best explained by appealing to a higher-level, domain-specific ToM module and not by appealing only to lower-level deficits.

G&S (2008, 132-133) discuss three case studies of very high-functioning individuals on the autism spectrum; two college students (a computer science student and a physics student) and a mathematics professor who had won the Fields Medal (Baron-Cohen et al. 1999). Baron-Cohen et al. (1999) administered three tests to the subjects: 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 presented with 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 (see Baron-Cohen et al. 1997; Baron-Cohen et al. 2001). Subjects performed well on the tests of executive function and folk physics, but they all exhibited deficits on the ToM test.

G&S (2008, 133) argue that since the ToM test in Baron-Cohen et al.'s (1999) study isolated the eye region, we can explain subjects' difficulty in identifying emotions by appeal only to a lower-level deficit:

> …all three had difficulty in inferring what someone was feeling, or paying attention to, from pictures of the eye region of the face […] indicating a problem with *lower-level domain-specific capacities* for face and gaze processing rather than MR [metarepresentation]. (2008, 133; emphasis added)

Notably, G&S do not discuss how development contributes to their explanation, confirming that lower-level deficits are doing the explanatory work in their account.[9]

There is a problem with G&S's interpretation of this study; they neglect to mention a relevant control task. To rule out the existence of some facial or gaze processing deficit, the subjects and the control group were asked to identify the gender of individuals 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 significantly 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, G&S (2008) do not address this aspect of the study, but one can imagine their potential response. 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 be due to a more domain-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 G&S argues that Baron-Cohen et al.'s (1999) control task does not allow us to determine whether the ToM deficits are due to the failure of a higher-level, domain-specific ToM module or of a lower-level, domain-specific face-related, emotion-

---

[9] Similarly, when they attempt to explain behavior in other studies that I will not discuss, G&S (2008, 131) appeal only to "impaired low-level domain-specific mechanisms," e.g., when explaining differences in performance between false-belief tests and false-photograph tests.

processing module. Recent studies in facial and auditory processing that I will discuss now lend support to the former 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 suggest that, though some lower-functioning persons with autism have difficulty with facial processing, in higher-functioning individuals the ToM deficit may be dissociated from facial-processing deficits. Lahaie et al (2006) used two tests to determine whether subjects' configural facial processing and response to the priming of facial parts differed significantly from the control group. The first test was a standard inverted-face design where participants matched upright and inverted faces. Face inversion is known to disrupt facial recognition because it inhibits normal configural processing of facial features.

In previous studies of the face inversion effect, subjects with autism performed better than controls in recognizing inverted faces but similarly or worse than controls with upright faces, causing some to argue that persons with autism have a deficit in configural facial processing (Lahaie 2006, 31). In Lahaie et al.'s (2001) study, subjects with autism were shown both faces with neutral expressions (to prevent a potential confound from faces showing emotion) and non-facial stimuli, called Greebles, that would also be inverted. Given previous studies, Lahaie et al. expected that subjects with autism would not show an inversion effect, supporting the deficiency in configural facial processing posited by previous studies (Lahaie et al. 2001, 31). Instead, both the control group and the subjects with autism exhibited the face inversion effect, which supports the view that higher-functioning persons with autism do not have a configural facial processing deficit. Joseph and Tanaka (2003) report a similar finding. Lahaie et al. (2001, 33) explain the discrepancy between their findings and previous studies'

findings by observing that previous studies attended only to subjects' error rates while Lahaie et al. attended also to concurrent response times by both groups, and previous studies were based on too small a number of trials (2001, 33).

Lahaie et al. (2006) designed the second test to determine whether 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 subjects with autism performed similarly to the control group on the priming task, suggesting that both groups benefited from being primed with configural information (Lahaie et al. 2006, 38). These findings suggest that individuals with autism do not have a configural facial processing deficit, 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.'s (2006, 38) study 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 subjects pay more attention to the eye region, but moreover after priming they processed the eye region *better* than other facial areas. Their overall performance at equal with the control group and their enhanced ability to process the eye region suggest not only that they lack 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 G&S's appeal to lower-level facial processing deficits to explain the ToM deficits in Baron-Cohen et al. (1999).

Humphreys et al. (2007) investigated whether higher-functioning persons with autism were able to recognize morphs of six basic facial expressions discussed by Ekman (1972). These morphs were computer-generated continua of facial expressions, e.g., from fear to surprise. Differences in emotion identification using these morphs were present between subjects with autism and the control group, with the most significant differences being in the recognition of

90% fear, 70% fear, and 70% disgust (Humphreys et al. 2007, 690). To determine whether subjects with autism differed due to a facial-processing deficit, a second test was administered to both groups where participants examined facial expressions and determined if two expressions were the same or different. Interestingly, there was no significant difference in the ability to make fine-grained comparisons between facial expressions between the two groups. This suggests that the difficulty discriminating emotions subjects with autism exhibited in Humphreys et al.'s first test was not due to a lower-level perceptual deficit.

Recent experiments on auditory processing by persons with autism provide further evidence for the modular ToM explanation and against an explanation appealing only to lower-level deficits (also see Golan et al. 2007). Rutherford et al. (2002) expanded upon Baron-Cohen et al.'s "reading the mind in the eyes" test by administering an auditory test designed to test ToM capabilities. In the "reading the mind in the voice" test, higher-functioning subjects with autism listened to a recording of someone speaking and were asked to describe the speaker's mental state (Rutherford et al. 2002, 190). Similar to the control test in the "reading the mind in the eyes" test, these subjects were asked to identify from two options the age range that best fit the age of the speaker. Subjects with autism were significantly impaired compared to controls in identifying the mental state of the speakers, but they did not differ significantly in assigning speakers to an age range. Rutherford et al. argue that these results suggest that the ToM deficit in higher-functioning persons with autism is amodal. This study is relevant to the present debate since the control task was designed to rule out lower-level auditory processing deficits, the sort of deficit to which G&S might appeal in a possible reply.

Kleinman et al. (2001) desired both to confirm the earlier "reading the mind in the eyes" experiments by Baron-Cohen et al. (1997; 2001) and to determine whether higher-functioning

persons with autism would exhibit similar deficits in an auditory task. Kleinman et al.'s (2001) control task required subjects to identify a speaker's gender from a recording. On the ToM task, subjects with autism performed significantly worse than the control group. On the control task, both groups scored above chance on the auditory task but neither group did so on the gender-identification visual task, which Kleinman et al. (2001, 35) take to indicate that this is a difficult task for anyone, i.e., a floor effect concern. That subjects with autism scored above chance on the auditory control task suggests that a lower-level auditory deficit is not responsible for their difficulty on the auditory ToM task.[10]

G&S could respond to these recent studies in two different ways. First, they could argue that though the studies discussed above seem to indicate that higher-functioning subjects with autism can have ToM deficits in the absence of lower-level visual or auditory processing deficits, this does not rule out the possibility of them having a higher-level deficit, e.g., 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. Neither of these responses works, so I argue that the best explanation of the ToM deficit appeals to a higher-level, domain-specific ToM module.

The first reply does not explain the behavior of the three subjects in Baron-Cohen et al. (1999). 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. G&S 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 recursive computations, as G&S admit (2008, 133). Furthermore, a study of a subject with damage to the left amygdala suggests that

---

[10] Further studies supporting this conclusion include Golan et al. (2006) and Golan et al. (2008).

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 executive function deficits (e.g., Griffith et al. 1999; Yerys et al. 2007).[11] Thus, it seems that G&S 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 would strongly detract from the force of G&S's argument. G&S (2008) emphasize that they are seeking a more parsimonious account of the ToM deficit in persons with autism by appealing to lower-level deficits. They emphasize their desire for parsimony with the cautionary phrase 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, G&S would need to posit a deficit in a lower-level module for emotion processing for vision and a deficit in a lower-level module for emotion processing for hearing (while all other visual and auditory processing modules are intact). Rather than explaining these deficits by adding a module for each modality, we can straightforwardly explain these studies by appealing to a single higher-level, domain-specific ToM module that accounts for deficits both in visual and auditory tasks.

5. Conclusion

In this paper, I have argued that the modular explanation of the ToM deficit in persons with autism is more plausible than a non-modular explanation, such as the putative explanation offered by G&S (2008). The plausibility of the modular account is supported by conceiving of

---

[11] For a review of the literature, see Hill (2004).

autism as a spectrum and by its ability to explain studies of higher-functioning individuals with autism.

References

Adams, Marcus P. 2009. "Empirical Evidence and the Knowledge-That/Knowledge-How Distinction." *Synthese* 170:97-114.

Adams, Marcus P. Forthcoming. "Modularity, Theory of Mind, and Autism Spectrum Disorder." *Philosophy of Science*.

American Psychiatric Association. 2000. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR)*. 4th ed., text revision. Washington, DC: American Psychiatric Association.

Ashwin, Chris, Simon Baron-Cohen, Sally Wheelwright, Michelle O'Riordan, and Edward T. Bullmore. 2007. "Differential Activation of the Amydala and the 'Social Brain' during Fearful Face-Processing in Asperger Syndrome." *Neuropsychologia* 45:2-14.

Baron-Cohen, Simon. 1989. "The Autistic Child's Theory of Mind: A Case of Specific Developmental Delay." *Journal of Child Psychology and Psychiatry* 30:285-297.

Baron-Cohen, Simon. 1998. "Does the Study of Autism Justify Minimalist Innate Modularity?" *Learning and Individual Differences* 10:179-191.

Baron-Cohen, Simon. 2009. "The Short Life of a Diagnosis." *New York Times* 9 November 2009.

> Available online at http://www.nytimes.com/2009/11/10/opinion/10baron-cohen.html.

Baron-Cohen, Simon, A. Leslie, and U. Frith. 1985. "Does the Autistic Child Have a 'Theory of

> Mind'?" *Cognition* 21:37-46.

Baron-Cohen, Simon, H. Ring, Sally Wheelwright, and T. Joliffe. 1997. "Is There a 'Language of

> the Eyes'? Evidence from Normal Adults and Adults with Autism or Asperger Syndrome."
>
> *Visual Cognition* 4:311-331.

Baron-Cohen, Simon, Sally Wheelwright, Valerie Stone, and Melissa Rutherford. 1999. "A

> Mathematician, a Physicist, and a Computer Scientist." *Neurocase* 5:475-483.

Baron-Cohen, Simon, Sally Wheelwright, Jacqueline Hill, Yogini Raste, and Ian Plumb. 2001.

> "The 'Reading the Mind in the Eyes' Test Revised Version: A Study with Normal Adults,
>
> and Adults with Asperger Syndrome or High-Functioning Autism." *Journal of Child*
>
> *Psychology and Psychiatry* 42:241-251.

Barrett, H. Clark and Robert Kurzban. 2006. "Modularity in Cognition: Framing the Debate."

> *Psychological Review* 113:628-647.

Bergeron, Vincent. 2007. "Anatomical and Functional Modularity in Cognitive Science: Shifting

> the Focus." *Philosophical Psychology* 20:175-195.

Bishop, D.V.M. 2000. "What's So Special about Asperger's Syndrome? The Need for Further

> Exploration at the Borderlands of Autism." In *Asperger Syndrome*, ed. Ami Klin, Fred R.
>
> Volkmar, and Sara S. Sparrow, 254-277. New York: The Guilford Press.

Caramazza, Alfonso. 1986. "On Drawing Inferences about the Structure of Normal Cognitive

> Systems from the Analysis of Impaired Performance: The Case for Single-Patient Studies."
>
> *Brain & Cognition* 5:41-66.

Caramazza, Alfonso and Michael McCloskey. 1988. "The Case for Single-Patient Studies."

    Cognitive Neuropsychology 5:517-525.

Carey, Susan. 2009. *The Origin of Concepts*. Oxford: Oxford University Press.

Carrington, Sarah J. and Anthony J. Bailey. 2009. "Are There Theory of Mind Regions in the

    Brain? A Review of the Neuroimaging Literature." *Human Brain Mapping* 30:2313-2335.

Carruthers, Peter. 2006a. *The Architecture of the Mind*. Oxford: Oxford University Press.

Carruthers, Peter. 2006b. "The Case for Massively Modular Models of Mind." In *Contemporary*

    Debates in Cognitive Science, ed. Robert J. Stainton, 3-21. Malden, MA: Blackwell

    Publishing.

Coltheart, Max. 1999. "Modularity and Cognition." *TRENDS in Cognitive Science* 3.3:115-120.

Coltheart, Max and Robyn Langdon. 1998. "Autism, Modularity, and Levels of Explanation in

    Cognitive Science." *Mind & Language* 13:138-152.

Cosmides, Leda & John Tooby. 1992. "Cognitive Adaptations for Social Exchange." In *The*

    Adapted Mind: Evolutionary Psychology and the Generation of Culture, ed. J. Barkow,

    Leda Cosmides, and John Tooby, 163-228. Oxford: Oxford University Press.

Ekman, Paul. 1972. "Universals and Cultural Differences in Facial Expressions of Emotion." In

    Nebraska Symposium on Motivation: vol. 1971, ed. J.K. Cole, 207-283. Lincoln, NE:

    University of Nebraska Press.

Factor, David C., Nancy L. Freeman, and Adam Kardash. 1989. "Brief Report: A Comparison of

    DSM-III and DSM-III-R Criteria for Autism." *Journal of Autism and Developmental*

    Disorders 19:637-640.

Fodor, Jerry. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.

Fine, C., J. Lumsden, and R.J.R. Blair. 2001. "Dissociation between 'Theory of Mind' and

    Executive Functions in a Patient with Early Left Amygdale." *Brain* 124:287-298.

Gerrans, Philip. 2002. "The Theory of Mind Module in Evolutionary Psychology." *Biology and*

    Philosophy 17:305-321.

Gerrans, Philip and Valerie E. Stone. 2008. "Generous or Parsimonious Cognitive Architecture?

    Cognitive Neuroscience and the Theory of Mind." *British Journal for the Philosophy of*

    Science 59:121-141.

Golan, Ofer, Simon Baron-Cohen, and Yael Golan. 2008. "The 'Reading the Mind in Films' Task

    [Child Version]: Complex Emotion and Mental State Recognition in Children with and

    without Autism Spectrum Conditions." *Journal of Autism and Developmental Disorders*

    38:1534-1541.

Golan, Ofer, Simon Baron-Cohen, and Jacqueline Hill. 2006. "The Cambridge Mindreading

    (CAM) Face-Voice Battery: Testing Complex Emotion Recognition in Adults with

    Asperger Syndrome." *Journal of Autism and Developmental Disorders* 36:169-183.

Golan, Ofer, Simon Baron-Cohen, Jacqueline Hill, and M.D. Rutherford. 2007. "The 'Reading the

    Mind in the Voice Test-Revised: A Study of Complex Emotion Recognition in Adults with

    and without Autism Spectrum Conditions." *Journal of Autism and Developmental Disorders*

    37:1096-1106.

Glymour, Clark. 2001. *The Mind's Arrows*. Cambridge, MA: MIT Press.

Griffith, Elizabeth, Bruce F. Pennington, Elizabeth A. Wehner, and Sally J. Rogers. 1999.

    "Executive Functions in Young Children with Autism." *Child Development* 70:817-832.

Hill, Elisabeth. 2004. "Evaluating the Theory of Executive Dysfunction in Autism."

    Developmental Review 24:189-233.

Humphreys, Kate, Nancy Minshew, Grace Lee Leonard, and Marlene Behrmann. 2007. "A Fine-Grained Analysis of Facial Expression Processing in High-Functioning Adults with Autism." *Neuropsychologia* 45:685-695.

Joseph, Robert M. and James Tanaka. 2003. "Holistic and Part-Based Face Recognition in Children with Autism. *Journal of Child Psychology and Psychiatry* 44:529-542.

Kamp-Becker, Inge, Judith Smidt, Mardjan Ghahreman, Monika Heinzel-Gutenbrunner, Katja Becker, and Helmut Remschmidt. 2010. "Categorical and Dimensional Structure of Autism Spectrum Disorders: The Nosological Validity of Asperger Syndrome." *Journal of Autism and Developmental Disorders* 40:921-929.

Kanner, Leo. 1943. "Autistic disturbances of affective contact." *Nervous Child* 2:214-250.

Karmiloff-Smith, Annette. 1998. "Development Itself is the Key to Understanding Developmental Disorders." *TRENDS in Cognitive Sciences* 2:389-398.

Karmiloff-Smith, Annett. 2009. "Nativism versus Neuroconstructivism: Rethinking the Study of Developmental Disorders." *Developmental Psychology* 45:56-63.

Kendell, Robert and Assen Jablensky. 2003. "Distinguishing between the Validity and Utility of Psychiatric Diagnoses." *American Journal of Psychiatry* 160:4-12.

Kleinman, Jamie, Paul L. Marciano, and Ruth L. Ault. 2001. "Advanced Theory of Mind in High-Functioning Adults with Autism." *Journal of Autism and Developmental Disorders* 31:29-36.

Klin, A., F.R. Volkmar, S.S. Sparrow, D.V. Cicchetti, and B.P. Rourke. 1995. "Validity and Neuropsychological Characterization of Asperger syndrome: Convergence with Nonverbal Learning Disabilities Syndrome." *Journal of Child Psychology and Psychiatry* 36:1127-1140.

Klin, Ami and Fred R. Volkmar. 2003. "Asperger Syndrome: Diagnosis and External Validity."

    Child and Adolescent Psychiatric Clinics 12:1-13.

Kupfer, David J., Michael B. First, and Darrel A. Regier. 2002. *A Research Agenda for DSM-V*.

    Washington, DC: American Psychiatric Association.

Lahaie, A., L. Motton, M. Arguin, C. Berthiaume, B. Jemel, and D. Saumier. 2006. "Face

    Perception in High-Functioning Autistic Adults: Evidence for Superior Processing of Face

    Parts, Not for a Configural Face-Processing Deficit." *Neuropsychology* 20:30-41.

Leslie, Alan M. 1987. "Pretense and Representation: The origins of 'Theory of Mind'."

    Psychological Review 94:412-426.

Machery, Edouard. Forthcoming. "Developmental Disorders and Cognitive Architecture." In

    Darwin and Psychiatry: Philosophical Perspectives, ed. A. De Block and P. Adriaens.

    Oxford: Oxford University Press.

Machery, Edouard & H. Clark Barrett. 2006. "Debunking Adapting Minds." *Philosophy of*

    Science 73:232-246.

Robertson, Lynn C., Robert T. Knight, Robert Rafal, and Arthur P. Shimamura. 1993. "Cognitive

    Neuropsychology is More Than Just Single-Case Studies." *Journal of Experimental*

    Psychology 19:710-717.

Robins, E., and S.B. Guze. 1970. "Establishment of Diagnostic Validity in Psychiatric Illness: Its

    Application to Schizophrenia." *American Journal of Psychiatry* 126:983–987

Rutherford, M.D., Simon Baron-Cohen, and Sally Wheelwright. 2002. "Reading the Mind in the

    Voice: A Study with Normal Adults and Adults with Asperger Syndrome and High

    Functioning Autism." *Journal of Autism and Developmental Disorders* 32:189-194.

Salmon, Wesley. 2006. *Four Decades of Scientific Explanation*. Pittsburgh, PA: University of

Pittsburgh Press.

Samuels, Richard. 1998. "Evolutionary Psychology and the Massive Modularity Hypothesis."

British Journal for the Philosophy of Science 49:575-602.

Scholl, Brian J. and Alan M. Leslie. 1999. "Modularity, Development, and 'Theory of Mind'."

Mind & Language 14:131-153.

Spirtes, Peter, Clark Glymour, and Richard Scheines. 2000. *Causation, Prediction, and Search*,

2nd ed. Cambridge, MA: MIT Press.

Stone, Valerie E. and Philip Gerrans. 2006a. "Does the Normal Brain Have a Theory of Mind?"

TRENDS in Cognitive Sciences 10:3-4.

Stone, Valerie E. and Philip Gerrans. 2006b. "What's Domain Specific about the Theory of

Mind?" *Social Neuroscience* 1:309-319.

Tanguay, Peter E., Julia Robertson, and Ann Derrick. 1998. "A Dimensional Classification of

Autism Spectrum Disorder by Social Communication Domains." *Journal of the American

Academy of Child and Adolescent Psychology* 37:271-277.

Tooby, John and Leda Cosmides. 1992. "The Psychological Foundations of Culture. In *The

Adapted Mind: Evolutionary Psychology and the Generation of Culture*, ed. J. Barkow,

Leda Cosmides, and John Tooby, 19-136. Oxford: Oxford University Press.

Volkmar, Fred R., Domenic V. Cicchetti, Donald J. Cohen, and Joel Bregman. 1992. "Brief

Report: Developmental Aspects of DSM-III-R Criteria Autism." *Journal of Autism and

Developmental Disorders* 22:657-662.

Volkmar, Fred R., Ami Klin, Bryna Siegel, Peter Szatmari, and 18 others. 1994. "Field Trial for

Autistic Disorder in DSM-IV." *American Journal of Psychiatry* 151:1361-1367.

Wing, Lorna. 1988. "The Continuum of Autistic Characteristics." In *Diagnosis and Assessment in Autism*, ed. Eric Schopler and Gary B. Mesibov, 91-110. New York: Plenum Press.

Yerys, Benjamin E., Susan L. Hepburn, Bruce F. Pennington, and Sally J. Rogers. 2007. "Executive Function in Preschoolers with Autism: Evidence Consistent with Secondary Deficit." *Journal of Autism and Developmental Disorders* 37:1068-1079.

Zaitchik, Deborah. 1990. "When Representations Conflict with Reality: The Preschooler's Problem with False Beliefs and 'False' Photographs." *Cognition* 35:41-68.