ɔity

ɔNITIVE SCIENCE

ɔnitive Science
)–54935–7 Hardback
)–54936–4 Paperback

s as they are published by placing a standing order.
e of difficulty, write to us at the address below with
series and one of the ISBNs quoted above.
llan Distribution Ltd, Houndmills, Basingstoke,

)F DAVID HUME

Neurofeminism

Issues at the Intersection of Feminist Theory and Cognitive Science

Edited by

Robyn Bluhm
Assistant Professor of Philosophy, Old Dominion University, USA

Anne Jaap Jacobson
Professor of Philosophy and Electrical and Computer Engineering, University of Houston, USA

and

Heidi Lene Maibom
Associate Professor of Philosophy, Carleton University, Canada

palgrave
macmillan

(2000) may be *too* broad. For instance, I may happen to feel very upset for no apparent reason while my friend, who has just been fired from his job, feels just fine. This does not empathy make.

2. Note, though, that the perspective taking that care theorists have in mind is often a great deal more onerous than that explored by social psychologists, as it does not simply involve trying to take up the perspective of another; it involves taking in the other in their alterity and feeling what they are feeling.
3. In the study at hand, the difference was not statistically significant.
4. Incidentally, one of the emotions that women were better than men at identifying in this study was sadness, contrasting with the findings of Erwin et al. (1992). This encapsulates the difficulty of the sex and gender difference research generally. The findings often fail to give a consistent picture of sex or gender differences.
5. On the perspective taking subscale of IRI (Interpersonal Reactivity Index) see Davis (1983).

4

The Role of Fetal Testosterone in the Development of the "Essential Difference" Between the Sexes: Some Essential Issues

Giordana Grossi and Cordelia Fine

1 Introduction

The Empathizing/Systemizing (E/S) hypothesis developed by Baron-Cohen and colleagues has two main goals: first, to explain the presence of brain, cognitive, and behavioral differences between the sexes; and second, to explain the pattern of symptoms associated with autistic syndromes. These two goals are connected, since Baron-Cohen argues that autism is the expression of an "extreme male brain" (e.g. Baron-Cohen 2002). Briefly, the E/S hypothesis proposes that levels of fetal testosterone (fT) influence brain development in such a way that lower levels of fT (more common in females) result in a 'female brain' that is "predominantly hard-wired for empathy" (Baron-Cohen 2003: 1). Empathizing is defined as "the drive to identify another's mental states and to respond to these with an appropriate emotion, in order to predict and to respond to the behavior of another person" (Baron-Cohen, Knickmeyer, and Belmonte 2005: 820). By contrast, higher levels of fT (more common in males) result in a 'male brain' that is "predominantly hard-wired for understanding and building systems" (Baron-Cohen 2003: 1). Systemizing is defined as "the drive to analyze a system in terms of the rules that govern the system, in order to predict the behavior of the system" (Baron-Cohen, Knickmeyer, and Belmonte 2005: 820). The "extreme male brain" of autism thus manifests as poor empathizing abilities twinned with superior systemizing abilities.

The social and political implications of the E/S hypothesis are readily apparent. While Baron-Cohen (e.g. Baron-Cohen 2003) is clear that a person of one sex may have the brain of the other (thus a woman may have a 'male' brain, and vice versa), on average the sexes will differ in their hardwired cognitive predispositions. This implies that gender inequalities are not due solely to gender discrimination or socialization, but at least partially to an "essential difference" (Baron-Cohen 2003), on average, between the sexes. (For an explicit argument of this kind with respect to sex ratios in math and physics, see Baron-Cohen 2007: 169.) Indeed, Baron-Cohen (2003: 185) argues that individuals with a 'female brain' are biologically predisposed toward occupations that, currently at least, are performed mostly by women, and thus "make the most wonderful counsellors, primary-school teachers, nurses, carers, therapists, social workers, mediators, group facilitators or personnel staff." By contrast, those with a 'male brain' supposedly enjoy a hardwired facility for traditionally male occupations, for example in science, engineering, business, law, and plumbing, all of which, it is argued, involve constructing and analyzing systems. The real-world implications of acceptance of the E/S hypothesis make it especially important that its assumptions, claims, and data are examined with care. In this chapter we focus on data and arguments regarding the relationship between fT, brain, and behavioral differences between the sexes. In the first section we evaluate the evidence for sex differences in Empathizing and Systemizing (herein E and S) abilities. In the second section we assess the arguments that fT organizes brain 'type.' We focus especially on the evidence presented by Baron-Cohen and colleagues in *Science* (Baron-Cohen, Knickmeyer, and Belmonte 2005), which we note is a prestigious journal with a wide circulation. In both sections we highlight numerous empirical, methodological, and conceptual inadequacies.

2 Sex differences in systemizing and empathizing

The E/S hypothesis asserts sex differences in E and S abilities. By way of support for this position, Baron-Cohen et al. (2005: 819) argue that:

> Although males and females do not differ in general intelligence, specific cognitive tasks reveal sex differences. Differences favoring males are seen on the mental rotation test (Shepard and Metzler 1971),[1] spatial navigation including map reading (Kimura 1999), targeting (Watson and Kimura 1991), and the embedded figures test (Witkin et al. 1962),[2] although there are conflicting studies regarding the latter (Hyde, Geiringer, and Yen 1975). Males are also more likely to play with mechanical toys as children (Hines, Allen, and Gorski 1992),[3] and as adults, they score higher on engineering and physics problems (Lawson, Baron-Cohen, and Wheelwright 2004). In contrast, females score higher on tests of emotion recognition (McClure 2000), social sensitivity (Baron-Cohen et al. 1999), and verbal fluency (Hyde and Linn 1988). They start to talk earlier than boys do (Fenson et al. 1994) and are more likely to play with dolls as children (Hines, Allen, and Gorski 1992). [For ease of reference, numbered citations in original have been replaced with author/date citations.]

Difficulties with these assertions fall into three types. First, in some cases it is contentious whether the tasks in which sex differences are observed fall under the purview of S or E ability. Second, the asserted sex difference may be under dispute, or contingent on social-contextual factors. Third, the potential role of experiential factors in the sex difference may be under-acknowledged.

2.1 Is it actually systemizing or empathizing?

We consider first the supposed greater male interest and skill in domains requiring systemizing. As noted earlier, systemizing is defined as the drive to analyze the rules that govern a system, with a view to predicting its behavior. This 'drive' is assumed to be content-free, that is, it applies to a variety of domains of systems, including technical, natural, abstract, social, organizable, and motoric systems (see Baron-Cohen 2002: 248). Systemizing interests would thus include grammar, physics, architecture, sociology, quilting, hair-dressing, and knitting; for some of which, we note, there is no evidence of greater male skill or interest.

Notwithstanding the imprecision of the definition, it is not immediately obvious to us that superior performance on mental rotation, spatial navigation, targeting, or embedded figures tasks constitutes convincing evidence of superiority in understanding and predicting systems. In fact, Baron-Cohen et al. (2005: 820) note that "it is unclear if the [embedded figures test] is really a test of systemizing or simply a test of good attention to detail". Unfortunately, this lack of clarity with regards to the specific cognitive operations involved in the cited tasks, or systemizing itself, creates rather too much leeway when it comes to empirical tests of the E/S hypothesis. For instance, Baron-Cohen has argued that the mental rotation test "involves systemizing because it is necessary to treat each feature in a display as a variable that can be transformed (e.g., rotated) and then predict the output, or how it

will appear after transformation" (Baron-Cohen 2007: 167, reference removed). Yet he and colleagues later questioned the validity of mental rotation as a systemizing measure, on the grounds that mental rotation ability does not correlate with proxies of fT exposure, stating that "mental rotation is not an ideal task for testing the elevated foetal testosterone (fT) hypothesis of [autism spectrum conditions]" (Knickmeyer et al. 2008: 995). Clearly, this is not a scientifically acceptable approach. If the hypothesis is that higher levels of fT create a more strongly systemizing brain, then cognitive tests should be defined *a priori* as systemizing or not. If a cognitive test that has been defined as systemizing fails to show an association with fT, then this constitutes lack of empirical support for the hypothesis.

Similar, although fewer, issues arise with regards to whether some of the skills cited by Baron-Cohen et al. (2005) as being superior in females actually constitute evidence of superiority in empathizing. Verbal fluency, for example – the ability to list as many words as possible from a particular category (like 'animals' or 'words beginning with the letter *p*') in a given period of time – bears no obvious link to empathizing ability. In addition, while it's certainly plausible that an ability to understand a caregiver's thoughts and intentions will facilitate language development, this doesn't imply that a relative delay in language development necessarily has poorer empathizing ability as its cause.[4] It is also not clear whether the superior performance of girls on the 'social sensitivity' task (Baron-Cohen et al. 1999, as cited in Baron-Cohen et al. 2005) is best explained in terms of superior female empathizing. This study found that while boys and girls aged seven, nine, and eleven years old were generally able to understand the erroneous mental states that led to faux pas, girls were more likely to identify when someone had said something she or he shouldn't have said. It's unclear, however, whether this difference arose because boys were less sensitive to the story characters' feelings, or whether they were simply more forgiving of the accidental and non-malicious hurting of others' feelings.

Exacerbating the imprecision of Baron-Cohen's approach in defining systemizing and empathizing is his use of self-report questionnaires – the Empathy Quotient (EQ, Baron-Cohen and Wheelwright 2004) and Systemizing Quotient (SQ, Baron-Cohen et al. 2003) – to measure E and S tendencies, or brain 'type.' This approach is problematic for two reasons. First, as Levy (2004: 322) has noted, the statements in the EQ and SQ are "often testing for the gender of the subject, by asking whether the subject is interested in activities which tend to be disproportionately associated with males or with females (cars, electrical wiring, computers

and other machines, sports and stock markets, on the one hand, and friendships and relationships, on the other)." The questionnaires are thus likely to make gender salient. Importantly, social psychological work has shown that priming gender increases self-stereotyping (e.g. Hogg and Turner 1987; James 1993; Steele and Ambady 2006; Chatard, Guimond, and Selimbegovic 2007). Indeed, even noting one's sex at the beginning of a questionnaire, as the EQ and SQ both require participants to do, can increase self-stereotyping (Sinclair, Hardin, and Lowery 2006). A serious concern, then, is that the responses on these self-report tests are significantly biased by gender-primed self-stereotyping.

Furthermore, self-report questionnaires do not measure actual behavior and often fail to predict behavior. For example, a now substantial literature shows that self-report measures of social sensitivity bear little relation to actual empathic accuracy. A review by Davis and Kraus (1997: 162) found that self-ratings of social sensitivity, empathy, femininity, and thoughtfulness had "minimal value" in the identification of good and poor social judges. More recent studies have also found only weak or non-significant correlations between self-estimates of ability and actual performance (Realo et al. 2003; Ames and Kammrath 2004; Voracek and Dressler 2006).

2.2 Is the sex difference real and reliable?

Two further issues arising from the evidence of sex differences cited by Baron-Cohen and colleagues are: first, some of these behavioral differences are under dispute; second, their existence is surprisingly sensitive to social-contextual factors. With regards to the first point, while there is support for a male advantage in mental rotation (e.g. Voyer, Voyer, and Bryden 1995), recent meta-analyzes have cast doubts on sex differences in spatial navigation and the embedded figure test (for a discussion of such literature, see Newcombe 2010). Similarly, the female advantages in verbal fluency and proficiency are not just of dubious relevance to the E/S hypothesis, but have also been questioned; differences may exist in children, but they tend to disappear with age (see Wallentin 2009). Female superiority in the cognitive component of empathizing (that is, inferring the thoughts and feelings of others) is also under question. Meta-analyzes have found superior decoding of nonverbal expressions of emotion in girls and women (Hall 1984; McClure 2000). However, research using the empathic accuracy test – a more realistic test of mindreading that assesses ability to infer a partner's thoughts and feelings in a genuine, unscripted social interaction – calls into question the assumption that females have an advantage in real-world mindreading

(Graham and Ickes 1997; Ickes 2003). Extensive use of this test has reliably found equivalent performance in the sexes (except in conditions to be discussed shortly; Ickes, Gesn, and Graham 2000).

With regards to the second issue, a growing body of social psychological research has demonstrated that sex differences can be significantly influenced by the social context in which the task is presented (for summary see Fine 2010a). Specifically, sex differences in performance are decreased or even eliminated when either the 'gendered' nature of the task, or the gender of the participant, is made less salient. Thus, sex differences in the performance of mental rotation tasks – the largest cognitive sex difference – have been significantly reduced and even eliminated altogether by simple changes such as presenting the task as associated with skill on 'feminine' compared with 'masculine' tasks (Sharps, Price, and Williams 1994), asking people to rotate stick figures rather than shapes (Alexander and Evardone 2008), or informing participants that women show superior performance (Moè 2009). A large literature on 'stereotype threat' (Steele 1997) has similarly shown significant effects of the social context on sex differences in math performance (for meta-analysis, see Nguyen and Ryan 2008). Briefly, stereotype threat refers to the detrimental effect on performance of a social context that highlights a relevant negative stereotype about one's social group (e.g. the stereotype that women are bad at math, during a math test). A recent meta-analysis of stereotype threat studies found that females, matched with males on real-world academic tests like the SAT, performed worse in math under stereotype threat (Walton and Spencer 2009). Moreover, the meta-analysis indicated that when stereotype threat was removed – generally by making gender seem less relevant to the task at hand – women actually outperformed their male peers who, from real-world tests, purportedly had the same mathematical ability.

Similar effects of task 'degendering' have been observed for both cognitive and affective components of empathizing. As noted earlier, research with the empathic accuracy test has reliably failed to find sex differences, regardless of whether the interacting dyads are strangers, friends, or romantic partners. However, when the test form was changed slightly to ask participants to rate the accuracy of their empathic judgments, female performance was enhanced (Ickes et al. 2000; Ickes 2003). Ickes (2003) suggested that this small change reminds women of the social expectation that women should be empathic. Similarly, Koenig and Eagly (2005) successfully closed the gender gap on a social sensitivity task by presenting it as a test of complex information processing.

Providing extra motivation to men to do well on empathizing tasks by offering social or financial incentives has also been successful in increasing male performance (Thomas and Maio 2008) and equalizing male/female performance (Klein and Hodges 2001), respectively. With regards to the affective component of empathizing (that is, experiencing an appropriate emotional response to another's mental state), Eisenberg and Lennon (1983) concluded from a meta-analysis that the female empathic advantage becomes vanishingly small as it becomes less and less obvious to the participant that something to do with empathy is being assessed. Thus, sex differences were greatest on self-report tests, smaller differences were seen when the purpose of the testing was less obvious, and few consistent differences were found for studies using unobtrusive physiological or facial/gestural measures as an index of empathy (although it should be noted that it's not clear how well such measures actually index affective empathy). Likewise, while Fabes and Eisenberg (1998) concluded that overall there is evidence of greater affective empathy in girls than boys, as with adults, this difference was smaller when based on observations rather than self-report or report by another (such as a parent).

The salience of participants' gender-identity also influences performance on gender-typed tasks. Thus females' mental rotation (McGlone and Aronson 2006) and math performance (Rydell, McConnell, and Beilock 2009) is improved, or becomes similar to that of males (Hausmann et al. 2009) when participants are primed to think of themselves in terms of a math-positive and/or non-gendered identity. Ryan, David, and Reynolds (2004) found that making a student- rather than gender-identity salient eliminated sex differences in care-based versus justice-based moral reasoning, and females asked to take the first-person perspective of a male character performed as poorly as males on emotion-knowledge tasks (Marx and Stapel 2006).

It is, we would suggest, problematic to attribute to differences between the 'female brain' and the 'male brain' sex differences in E and S that can be so readily reduced and even eliminated by simple social manipulations that diminish the salience of stereotypical expectations.

2.3 The purported 'innateness' of sex differences

Finally, sex differences in empathizing and systemizing abilities and interests, when present, might stem from experiential factors. While Baron-Cohen acknowledges that culture plays a "major role," he regards gender socialization factors as "amplifying... partly innate differences" (Baron-Cohen 2007: 169). Baron-Cohen et al. (2005: 819–820) cite three

lines of research as evidence that there is a "biological foundation" to purported sex differences:

> Male rats perform significantly better than females do on the radial arm and Morris water maze (Roof et al. 1993). This sex difference is eliminated by castrating males or by treating females with testosterone neonatally (De Vries and Simerly 2002). Human males also commit fewer errors and require less time to complete a 'virtual' maze (Moffat, Hampson, and Lee 1998). Young male vervet monkeys prefer to play with toy trucks, whereas young female vervets prefer dolls (Alexander and Hines 2002). This finding suggests that sex differences in toy preferences in children result, in part, from innate biological differences. Biological contributions to social interest are suggested by studies of human infants. When one-day-old babies are presented with either a live face or a mechanical mobile, girls spend more time looking at the face, whereas boys prefer the mechanical object (Connellan et al. 2001).[5] [For ease of reference, numbered citations in original have been replaced with author/date citations.]

Examination of these three lines of evidence – from maze performance in rats and humans, toy preferences in monkeys, and newborn preferences for mechanical versus social stimuli – in each case yields conceptual and empirical difficulties, which we discuss in turn.

2.3.1 *Maze performance in rats and humans*

First, Baron-Cohen et al. cite data from non-human animals – rats – as evidence that similar sex differences in spatial navigation in humans are biologically inherent. We begin by noting that the study cited as evidence of superior male rat performance on the radial arm and Morris water maze task (Roof et al. 1993) was a lesion study that found no sex differences in neurologically intact animals.[6] Moreover, DeVries and Simerly (2002) do not mention studies of spatial skills in the rat. Moffat et al. (1998) was an MRI investigation of planum temporale and corpus callosum morphology in left handers, and did not involve a virtual maze task. Baron-Cohen et al. may instead be referring to Moffat, Hampson, and Hatzipentalis (1998), who found sex differences in a task requiring participants to navigate a virtual maze. However, we note that participants were asked to fill out a demographic questionnaire before performing the behavioral test. It is therefore plausible that behavior was influenced by stereotype threat.

Moreover, no justification is provided for selecting rats as an appropriate comparison with humans. Humans are cognitively and neurologically dissimilar to rats in potentially important ways (see Hines 2004: 215), and it is not known whether the same mechanisms are involved in spatial navigation in the two species. Underlining the need for caution in extrapolating from rats from humans is research showing that it is impossible to generalize even within the same species. For example, in a meta-analysis of spatial behavior in rodents, Jonasson (2005) found a sex difference favoring male rats in two different types of mazes (water and radial arm), but the difference varied depending on the strain of rats. Importantly, the difference disappeared, or was reversed, in mice.

2.3.2 *Toy preferences in monkeys*

Baron-Cohen et al. (2005: 820) next cite an observational study of vervet monkeys' toy play behavior (Alexander and Hines 2002) as evidence that human sex differences in children's play behavior is due in part to "innate biological differences." This study compared contact time with 'masculine' toys (a ball and a police car), 'feminine' toys (a toy pan and a doll), and 'neutral' toys (a picture book and a stuffed dog), presented serially in the vervet enclosure. Between-sex contrasts showed greater male interest in the 'masculine' toys, and greater female interest in the 'feminine' toys. The sexes showed equal interest in the 'neutral' toys. Within-sex contrasts found only that females had greater percentage contact with 'feminine' toys than with 'masculine' toys. More recently, Hassett et al. (2008) ran a similar study with male and female rhesus monkeys, in which they compared interaction (using two dependent variables, total frequency and total duration of contact) with 'masculine' wheeled toys versus 'feminine' stuffed toys. Between-sex contrasts found that males and females were equally interested in the wheeled toys. Males and females also spent a similar duration of time with the stuffed toys, but females had a greater total frequency of interaction with these toys. Within-sex contrasts revealed that males preferred wheeled toys over stuffed toys, while females showed no preference.

There are two important points to be made about these findings (see Fine 2010a and for a further critique of the earlier study, see also Jordan-Young 2010). First, there are issues regarding the choice of 'feminine' toys. Although in human culture cooking utensils are associated with females due to their role in domestic caretaking, it is entirely unclear why a female predisposition toward a toy pan should be anticipated in monkey populations, which do not enjoy the art of heated cuisine. The pan was, however, the most popular toy for female

Alexander and Hines (2002) suggested that this may have been due to the appeal of its red color. This raises the possibility that other sex differences in toy preferences may have been due to confounding factors. It is also worth noting that a stuffed animal (a dog) was used as a 'neutral' stimulus by Alexander and Hines (and was in fact the most popular toy with male vervets), yet stuffed animals were used as 'feminine' stimuli by Hassett et al. even though, as Hines and Alexander (2008) point out, boys and girls like stuffed animals equally.

Second, the results from the two studies are somewhat contradictory with each other. Alexander and Hines (2002) found greater male than female interest in 'masculine' toys, while Hassett et al. (2008) found that males and females were equally interested in 'masculine' toys. Moreover, while Hassett et al. found that rhesus males spent significantly more time with 'masculine' toys than with 'feminine' ones, Alexander and Hines found no such preference in vervets. Likewise, Hassett et al.'s observation of greater female than male interest in stuffed animals (significant for only one dependent measure) was not observed by Alexander and Hines. In other words, the studies together leave some uncertainty regarding the reliability of both between- and within-sex differences in gendered toy preferences.

2.3.3 Newborn preferences for mechanical versus social stimuli

As a final line of evidence for the 'innateness' of sex differences in empathizing/systemizing tendencies, Baron-Cohen et al. (2005) refer to a study of newborn looking preferences which compared neonates' looking time at a live face versus a mobile (Connellan et al. 2000). Male and female babies spent equal amounts of time looking at the face (that of the first author): both sexes, on average, spent just under half the total presentation time (approximately a minute) looking at Connellan's face.[7] However, males looked longer at the mobile than did females (51.9 percent of presentation time versus 40.6 percent for females) and females looked longer at the face than the mobile (49.4 percent versus 40.6 percent of presentation time). A detailed critique of the methodological flaws of this study has been provided by Nash and Grossi (2007). These flaws range from the many differences between the stimuli (each of which could have been responsible for the observed differences), to experimenter expectancy effects (the first author was herself the live face, and controlled the movement of the mobile), to the authors' non-standard procedure for measuring looking time preference (serial presentation rather than the simultaneous presentation methodology that is standard in infant visual preference research). These serious methodological issues render questionable the study's contribution to the scientific literature; notably, the findings have not been replicated. Moreover, no attempt is made by the authors to justify the assumption that newborn looking preferences anticipate future propensity for complex, culture-dependent skills such as physics or empathizing (Levy 2004; Nash and Grossi 2007).

2.4 Summary of evidence for sex differences in E and S

The claim for the existence of strong and reliable sex differences in E and S is not supported by the extant evidence. The imprecision in defining E and S (as 'drives,' not as sets of specific cognitive abilities) has made it difficult to find cognitive tasks that convincingly capture these constructs; furthermore, when sex differences are present they are not reliable and can be explained by other mechanisms. The strongest evidence for sex differences in E and S comes from self-report questionnaires which are, as discussed, fraught with methodological problems. The argument that some of these differences have a biological foundation because they are observed in other animals and in newborns is also not convincing. The choice of the species used to support data on humans is generally not properly justified. Moreover, data from monkey and newborn studies are ambiguous or have not been replicated. By contrast, the role of gender socialization processes in gendered preferences is well documented (e.g. Bussey and Bandura 1999; Martin and Ruble 2004; Miller, Trautner, and Ruble 2006; Leaper and Friedman 2007).

3 Fetal testosterone organizes functional brain 'type'

The E/S hypothesis predicts that differences in fT levels are associated with differences in specific brain structures that are, in turn, associated with differences in cognitive style or behavior. Our discussion of this body of evidence falls into three parts. We begin by discussing the conceptual roots of this aspect of the E/S hypothesis and the assumptions about the nature of the developmental process implicit in such an account. Next, we evaluate the strength of empirical support for the prediction that fT levels correlate with later E and S abilities. Third, we discuss the evidence regarding sex differences in the brain in relation to the prediction that fT influences brain structure in ways that have functional implications for E and S ability.

3.1 The conceptual roots of the E/S hypothesis

The conceptual roots of the E/S hypothesis lie in what has been termed the "orthodox view of brain sexual differentiation" (Breedlove, Cooke, and Jordan 1999). Briefly, this view holds that the gene-directed development of testes in the male fetus results in high levels of gonadal fT that direct the development of male genitalia, and 'organize' a male brain that produces male behavior (in some cases after these brain structures are 'activated' by circulating sex hormones in pubescence and adulthood). The organizational-activational hypothesis, first proposed by Phoenix et al. (1959), seeks to account for sexually dimorphic behavior – in particular, behavior tied to reproduction (for example, frequency of mounting and lordosis) – and empirical tests of the hypothesis are usually conducted with non-primate mammals. The hormonal environment is manipulated during the critical period that masculinization of the external genitalia takes place, and the effects on brain organization and/or sexually dimorphic behavior are observed (for a brief summary, see Breedlove et al. 1999). (The timing of the critical period varies across species; for example, in rats the critical period for sexual differentiation includes the early postnatal period. Thus for simplicity, we sometimes use the term 'early' to encompass both prenatal and neonatal testosterone levels.)

In a highly influential proposal, Geschwind and colleagues extrapolated such accounts of brain sexual differentiation in animals to humans. Research by Diamond and colleagues showed that high levels of neonatal testosterone in male rats were associated with a relatively thicker right hemisphere cortex: in castrated male rats, the normal right-thicker-than-left cortical asymmetry was reversed (see Diamond 1991). Partly on the basis of such findings in rats, Geschwind and Behan (1982) proposed that similarly in humans the higher levels of fT in males stimulates growth of the right hemisphere while inhibiting growth of the left. Further, Geschwind and Galaburda (1987) proposed cognitive sequelae to these differences in brain development, suggesting that these structural differences enable greater male facility for visuospatial and mathematical processing, but delay language development, compared with females.

Baron-Cohen and colleagues cite Geschwind's work as supportive of a role for fT in sexual differentiation of the brain (e.g. Lutchmaya, Baron-Cohen, and Raggatt 2002a; Chapman et al. 2006). However, numerous differences in the way that early hormones affect rodents and humans have led to the conclusion that the "dominant rat and mouse models of sexual differentiation seem unlikely to apply to human sexual differentiation" (Wallen 2005: 8). Hines (2004), moreover, has noted that rat brains differ to human brains in important ways, with proportionally more area devoted to sensory functions and fewer association areas devoted to more complex, higher-order cognition. Thus, it cannot be assumed that the brain changes wrought by early testosterone in rats would be preserved in humans. In line with these concerns, fT does not appear to have the same effect on right hemisphere growth in humans as it does in rats. Neither a large post-mortem study of fetal brains (Chi, Dooling, and Gilles 1977; see discussion in Bleier 1986) nor a structural neuroimaging study of 74 newborns found evidence for a relatively larger right hemisphere in human males (Gilmore et al. 2007). Moreover, as Nash and Grossi (2007) note, post-mortem and neuroimaging studies of adult brains have also failed to demonstrate a relatively larger right hemisphere in males, either overall, or specifically in the parts of the brain thought to be involved in spatial processing.

Beyond the difficulty of extrapolating from non-primate mammals, it's also unclear whether the orthodox view provides an adequate account of brain sexual differentiation even in these animals. It has been noted that establishing a simple causal pathway from early T to structural brain differences to behavioral differences in mammals has so far proved impossible at a higher level than the brainstem (see, for example, brief discussions in Breedlove et al. 1999; De Vries 2004). This failure may be explained by research showing that the effects of early testosterone on brain and behavior may be indirect. For instance, Moore et al. (1992) have shown that the higher level of neonatal testosterone in male rat pups produces odor cues (in the pup urine) that elicit differential treatment from the mother (greater anogenital licking), and that this maternal behavior contributes to brain and behavioral differences between the sexes. In other words, early T may affect brain and behavior indirectly, via social environmental effects.

As Moore (2002) points out, the orthodox view of brain sexual differentiation is premised on a 'development to' perspective, according to which the environment merely influences the individual's progress 'to' a genetically encoded phenotype, via gene-directed effects on hormones and, thus, the brain. Such a perspective is implicit in Baron-Cohen's writing, in which he refers to socialization factors influencing (for example, "amplifying"; Baron-Cohen 2007: 169) what is innately or biologically specified (see Levy 2004: 323). The 'development to' perspective is predominant in psychological science (although see, for example, Karmiloff-Smith 2007, and Westermann et al. 2007, for overviews of cognitive neurodevelopmental approaches that explicitly reject

a 'development to' perspective). However, as Lickliter and Honeycutt (2003: 819) have noted, in other areas of developmental science a "conceptual revolution" has led to a 'development from' perspective, according to which there is no pre-specified developmental pathway. Rather, every developmental step, including even behaviors previously considered instinctual or innate (e.g. Blumberg 2005 – see references to Gottlieb's research), is constructed from the complex and dynamic interaction between environmental stimuli, genotype, and the organization of the nervous system in a particular developmental stage.

As Moore (2002: 65) notes, research strategies that observe only early hormones and behavioral outcomes leave "lots of unexplored territory and many possible pathways, perhaps convoluted ones, from the early hormones and end points of interest." Indeed, given the complexity of the role of early hormones in the developmental process that is becoming apparent even in non-primate animals, it is not clear whether it is even plausible to predict that in humans fT levels (independently of sex) will predict later neurological or behavioral outcomes. With this in mind, we turn to the empirical evidence for associations between fT, brain structure, and E/S profile.

3.2 fT and E/S profile in clinical and non-clinical populations

Baron-Cohen et al. (2005: 822) cite two types of evidence as support for the idea that fT levels correlate with E/S profile:[8]

> In humans, exposure to atypically high levels of prenatal androgens results in masculine behavior and ability patterns (Berenbaum 2001). For example, females with congenital adrenal hyperplasia (CAH), a genetic condition that elevates fetal testosterone (FT), show tomboy behavior (Hines and Kaufman 1994). Normal interindividual variation in prenatal hormone levels, measured in amniotic fluid, correlates with later sex-typed behavior (Grimshaw, Sitarenios, and Finegan 1995; Lutchmaya, Baron-Cohen, and Raggat 2002a, 2002b; Knickmeyer et al. 2005a) [Numbered citations replaced with author/date citations for ease of reference.]

In the following two sections we examine data from females with CAH, and then studies seeking to establish correlations between indices of fT levels and later sex-typed behavior.

3.2.1 Girls with CAH

There are a number of important observations to make with regards to the implications of the behavior of girls with CAH for the E/S

hypothesis. First, the demonstration of *any* male-typical (or 'tomboy') behavior in girls with CAH does not constitute support for the E/S hypothesis: rather, the behavior must demonstrate stronger systemizing and/or lesser empathizing tendencies. The study cited as showing tomboy behavior in girls with CAH (Hines and Kaufman 1994) observed rough-and-tumble play, and asked children to report the sex of their three preferred playmates. It is not clear that either measure relates to empathizing versus systemizing, and although Baron-Cohen (2007) has argued that rough-and-tumble play may reflect males' lower levels of empathy, it is worth noting that successful rough-and-tumble play is likely to demand quite high sensitivity to cues from one's play partner. The relevance of Hines and Kaufman's study to the E/S hypothesis is therefore unclear. Moreover, whereas CAH girls tended to report a preference for boys as playmates when compared to control girls (44.1 percent vs. 11.2 percent), there were no differences between CAH girls and control girls for rough-and-tumble play.

Second, there is a difficulty in interpreting studies showing an enhanced preference of girls with CAH for male-typical activities (for a comprehensive review of these data, see Jordan-Young 2010). This is because such research has made no attempt to investigate whether girls with CAH are drawn to some intrinsic quality in boyish toys and activities, or whether they are drawn to them simply by virtue of the fact that they are associated with males (Bleier 1986; Fine 2010a; and for general discussion of issues with measurement of sex-typed interests, and arguments regarding the potential psychological effects of the intrusive medical management and social expectations experienced by this clinical group, see Jordan-Young 2010). For example, girls with CAH score more similarly to boys than do unaffected female controls on the Pre-School Activities Inventory (PSAI, Golombok and Rust 1993) which taps interest in traditionally feminine toys and activities, including jewelry, pretty things, dressing up in girlish clothes, and pretending to be a female character (Hines et al. 2003). Likewise, women with CAH asked to recall their childhood activities responded significantly differently from controls on a questionnaire that, among other items, asks about use of cosmetics and jewelry, hating feminine clothes, the gender of admired or imitated characters on TV or in movies, and whether they dressed up more as male or female characters (Meyer-Bahlburg et al. 2006). A study of children with a condition causing either partial or complete androgen insensitivity (46,XY karyotype) found that greater prenatal androgen exposure led to less interest in activities like ballet, gymnastics, playing hairdresser, working with clay, and dressing up as a fairy, a witch, or a woman, but more interest in basketball, playing

spaceman, and dressing up as an alien, a cowboy, a man, or a pirate (Jürgensen et al. 2007). It is unclear to us what form of brain masculinization could lead to a cognitive predisposition for dressing up as an alien rather than a witch, lack of interest in jewelry and cosmetics, or masculine costumes over feminine ones.

Studies of toy preferences suffer the same confound. It could be argued that male-typical toys such as vehicles and construction toys facilitate systemizing more than do female-typical toys such as dolls and tea sets, which may offer more opportunity for empathizing. However, it is not clear why differences between girls with and without CAH (or indeed sex differences) are not seen for gender-neutral toys like puzzles and sketchpads, which would also appear to facilitate systemizing more than empathizing (Fine 2010a). As Bleier (1986: 150) pointed out in her critique of early studies in this area, "authors and subsequent scientists accept at face value the idea of tomboyism [such as play preferences, clothing preferences, career interests, and so on] as an index of a characteristic called *masculinity*, presumed to be as objective and innate a human feature as height and eye color. Yet 'masculinity' is a gender characteristic and, as such, culturally, not biologically, constructed." By failing to specify *a priori* what properties of toys or activities will be differentially appealing to boys and girls, researchers interested in the hormonal origins of gendered toy preferences can simply replace toys that fail to elicit the expected sex difference in preference (Pasterski et al. 2005: see 269). So while there is evidence that girls with CAH are drawn to 'masculine' activities and toys, relative to unaffected controls, so far researchers have failed to adapt their methodologies in response to Bleier's criticism.

There has also been research interest in the question of whether girls with CAH show superior mental rotation abilities. Although a recent meta-analysis suggests that girls with CAH have enhanced mental rotation skills (Puts et al. 2008; although see Jordan-Young 2010: 304 for a criticism of this conclusion), this may be a consequence of their greater male-typical play. Videogames enhance spatial skills (Dorval and Pépin 1986; Feng, Spence, and Pratt 2007; Cherney 2008) and there is evidence suggestive that play behavior likely to be differentially experienced by the sexes may also enhance spatial skills (Sprafkin et al. 1983; Baenninger and Newcombe 1989; Levine et al. 2005). Recent studies have also investigated personality traits in girls with CAH. However, so far the findings have been somewhat inconsistent. Thus, older girls and women with CAH report less social skills, tender-mindedness, and interest in infants than unaffected relative controls (Knickmeyer et al. 2006a; Mathews et al. 2009). However, these studies found no differences in social communication ability and dominance (which includes traits such as aggression, authoritativeness, and competitiveness; for a rejection of increased aggression associated with CAH, see also Jordan-Young 2010). Moreover, self-report measures may correspond poorly to actual behavior (as noted in section 2.1), and maternal reports may be biased by knowledge of the child's clinical status.

3.2.2 fT and E/S profile

The second category of evidence cited by Baron-Cohen and colleagues refers to studies that try to establish a link between fT exposure and "ability patterns" (Baron-Cohen, Knickmeyer, and Belmonte 2005: 822). There are two important points worth making about this empirical approach. First, as Hines (2004) has argued, a role for prenatal testosterone can only be expected for behaviors for which there are genuine differences between the sexes. As discussed earlier (see section 2.2), in humans some purported sex differences are under dispute, or contingent on social-contextual factors. Second, while both amniotic testosterone, or aT (sampled from the amniotic fluid during amniocentesis), and maternal testosterone, or mT (sampled from the mother's blood), have been used as proxies for fT exposure, there is currently no satisfactory evidence that either is related to actual fT exposure. In their review of this issue, van de Beek et al. (2004) suggest aT as the best index of fT exposure, but they also acknowledge the lack of much understanding of the relationship between levels of T in the amniotic fluid – the main source of which is fetal urine – and in the fetal blood. Indeed, as Knickmeyer et al. (2005b: 521) acknowledge, "there is no direct evidence to either support or contradict" the assumption that aT is correlated with the levels of testosterone acting on the fetal brain. Likewise, the relationship between maternal T and fetal levels is unclear. One clinical study that measured fT directly did find that it correlated with maternal T (Gitau, Adams et al. 2005). However, maternal T levels are not higher in women carrying boys than in those carrying girls, which suggests that "maternal serum androgen levels are not a clear reflection of the actual exposure of the fetus to these hormones" (van de Beek et al. 2004: 664).

It is a cause of concern that claims may be made about the prenatal, hormonal origins of sex differences on the basis of supposed biological markers of fT exposure that, remarkably, have unknown relationships with actual fT exposure. Despite this, a growing number of studies have investigated relationships between aT or maternal T and later cognitive

or social abilities. In terms of evaluating these studies, it is worth noting that four criteria should be fulfilled in order to claim support for the E/S hypothesis. First, the dependent variable(s) should be plausibly characterizable as part of an E or S skill set. Second, the dependent variable should be methodologically soundly measured. Third, the dependent variable should show a reliable sex difference in the predicted direction. Finally, correlations between the fT proxy and the dependent variable should be seen within the sexes, as well as in the group as a whole (otherwise fT may be confounded with the effects of gender socialization).

Table 4.1 summarizes the data from all such studies. We note that in not a single study are all these criteria satisfied. A detailed critique of each study can be found in the Appendix, where we briefly discuss first, in chronological order, the studies that relate aT (or mT) to purported measures of empathizing, followed by studies concerned with systemizing, then gender-typical play behavior.[9] Based on the analysis of this literature, the evidence concerning "[n]ormal interindividual variation" (Baron-Cohen, Knickmeyer, and Belmonte 2005: 822), both that provided by Baron-Cohen and colleagues as well as other relevant work, yields a scattered and inconsistent picture. (Concern has also been expressed regarding inconsistencies in the treatment of statistical outliers and statistical modeling procedures in the analysis of aT data by Baron-Cohen and colleagues (Jordan-Young 2010: see 219 and endnote).) Regularly, behaviors are tested that appear to assess something other than E or S ability, methodologies are often weak and sample sizes inadequate, behavioral differences on performance tests are not reliably and consistently observed, and functions relating fT and behavior are often different from the predicted one, present only in one sex, explained by sex, or completely absent.

3.3 Sex differences in the brain

The E/S hypothesis predicts a relationship between fT level and brain structure, and requires that these structural differences have implications for E/S function. To our knowledge, no research has investigated relationships between fT and brain structure in humans. We therefore focus here on the prediction that there are functionally significant sex differences in the brain. By way of support for this aspect of the hypothesis, Baron-Cohen et al. (2005: 820) begin by referring to sex differences in brain size (citing Giedd et al. 1996), "a difference that is driven more by white matter than by gray" (citing Allen et al. 2003, and Lüders et al. 2005), relatively larger female corpus collosum size (citing Allen et al. 2003), and larger amygdala volume in boys (citing

Caviness et al. 1996) and possibly also men (citing Goldstein et al. 2001),[10] greater numbers of neurons in the male cerebral cortex (citing Pakkenberg and Gundersen 1997), that are more densely packed (citing Rabinowicz et al. 2002) although with exceptions in certain regions (citing Witelson, Glezer, and Kigar 1995). They go on to suggest that some of these structural differences "indirectly suggest a pattern of increased local connectivity and decreased interhemispheric (or long-range) connectivity in the male brain" (2005: 820). They then argue that studies showing more bilateral activation in females during language-related tasks (Shaywitz et al. 1995; Baxter et al. 2003), and a magnetoencephalography (MEG) study of grocery choices made by eight men and eight women in a shopping simulation (Braeutigam et al. 2004) reporting "increased phase locking between frontal and parietal sites in women", also suggest a stronger skew toward local connectivity in males.

Before turning to the question of what implications, if any, such structural differences have for function, there are two important points to be made about the structural claims themselves. First, on average, males have larger brains than females, and there are currently unresolved questions regarding whether structural differences in the brain are due to size rather than sex *per se*. Thus it has been argued that brain size, not sex, is the main variable affecting ratios of gray to white matter volumes (e.g. Lüders, Steinmetz, et al. 2002; Im et al. 2008). Recently, Lüders et al. (2009) found that the ratios of gray and white matter, relative to total brain volume, did not differ between men and women matched for brain size. (They did, however, find some regional volume differences in their matched groups, with larger gray matter volumes in women than men.)

A second issue is that isolated findings of particular sex differences in the brain may be spurious. A particular problem for sex differences research arises from the practice of testing for sex differences by default. As Kaiser et al. (2009: 54) have noted, classifying by sex is a "natural default" and "seemingly effortless and obvious in brain research." The concern is that false positive results arising from sex comparisons are reported, while true negatives are not (Maccoby and Jacklin 1974; for discussion specifically in relation to neuroimaging research, see Fine 2010a, 2010b). That spurious results can lead to a misleading impression of the viability of a hypothesized sex difference has been well demonstrated by two purported sex differences cited by Baron-Cohen et al. (2005): in the corpus callosum, and in the degree of lateralization of language function. A meta-analysis of 49 post-mortem and MRI investigations of the corpus callosum concluded that, even controlling for

Table 4.1 Summary of studies investigating aT/ mT and E/S ability

Dependent variable	Purported E/S measure	Study	Behavioral sex difference	Relationship with aT[a]	Relationship with aT within sexes
Vocabulary[b]	E	Lutchmaya et al. (2002b) (volume 24: 418–424)	Yes (higher for girls)	Yes	No
Eye contact	E	Lutchamaya et al. (2002a) (volume 25: 327–335)	Yes (higher for girls)	Yes (linear and quadratic)	None for girls; linear and quadratic for boys
Social relationship skills[b]	E	Knickmeyer et al. (2005a)	A trend favoring girls	Yes (negative)	No
Mental, affective, and intentional terms in triangle movie	E	Knickmeyer et al. (2006b)	Yes for affective (higher for girls) and neutral (higher for boys) terms	No for mental or affective terms, yes for intentional (negative) and neutral (positive) terms	No for mental, affective, and neutral terms. Intentional terms: none for girls, trend for boys
EQ[b]	E	Chapman et al. (2006)	Yes (higher for girls)	Yes (negative)	Negative for boys; none for girls
Reading the Mind in the Eyes test	E	Chapman et al. (2006)	No	Yes (negative)	Negative for both boys and girls
Block building task and embedded figures[d]	S	Finnegan et al. (1992)	No	Analyzes were not carried out	Negative for block building in girls, none in boys. None for embedded figures
Mental rotation	S	Grimshaw et al. (1995)	No differences in accuracy	No	Positive for girls who used a rotating strategy; negative for boys[e]
Restricted interests[b]	S	Knickmeyer et al. (2005a)	Yes (boys had more restricted interests than girls)	Yes (positive)	None for girls, positive for boys
SQ-C[b]	S	Auyeung et al. (2006)	Yes (boys scored higher than girls)	Yes (positive)	Positive for both sexes[f]
Block design	S	Auyeung et al. (2009a)	No	No	No
Pre-school Activities Inventory (PSAI)[b] (measure of sex-typical behavior)		Hines et al. (2002)[c]	Yes (boys scored higher than girls)	Analyzes were not carried out	Positive for girls, none for boys
Gender-typical play[b]		Knickmeyer et al. (2005b)	Yes (girls scored higher on femininity scales and boys scored higher on masculinity scales)	No	No
Gender-typical play		van de Beek et al. (2009)	Yes (preference for specific toys)	No	No
Pre-school Activities Inventory (PSAI)[b]		Auyeung et al. (2009b)	Yes (boys scored higher than girls)	Yes (positive)	Yes (positive for both sexes)

Notes:
a = data from both sexes pooled together; an aT effect can be explained by sex if sex is not removed from the analyzes.
b = data were based on mothers' report.
c = maternal T.
d = the authors employed a variety of cognitive tests in their study, some of which are difficult to summarize. Here we limit our analysis to tasks that have been linked to systemizing. See the Appendix for more information.
e = the correlation became significant when two data points were removed, quite arbitrarily, from the analyzes.
f = beta values for girls have opposite signs in the text and table 3 in the original article.

overall brain size, there is no reliable sex difference in the size or shape of this structure (Bishop and Wahlsten 1997). Bishop and Wahlsten particularly note the issue of spurious results arising from small sample sizes. So while Baron-Cohen et al. acknowledge Bishop and Wahlsten's findings, it's not clear how much weight the study of 23 men and 23 women (Allen et al. 2003), cited by Baron-Cohen et al. (2005) as evidence of greater corpus callosum size in females, should be credited. Similarly, while the idea of greater male lateralization of language function has enjoyed considerable popularity (for critique of the hypothesis prior to investigation using neuroimaging technologies, see Bleier 1986), recent meta-analyzes of functional neuroimaging lateralization studies suggest that linguistic functions are not more bilateral in women compared to men (Sommer et al. 2004, 2008). Sommer and colleagues' (2008) meta-analysis of dichotic listening tasks also failed to find evidence of lateralization differences between the sexes. (If males are more lateralized for language, they should have a stronger "right ear advantage" than females for language presented to the left, language-dominant hemisphere via the right ear.) Moreover, while there have been suggestions that males are more likely to suffer aphasia following stroke damage to the left hemisphere (which would be consistent with the idea that females' greater right hemisphere language function would serve a protective effect), as Wallentin (2009) notes, the Copenhagen aphasia study of more than 1000 patients found no effect of sex or side of stroke lesion (Pedersen, Vinter, and Olsen 2004; cited in Wallentin 2009). Similarly, if linguistic functions were less lateralized in women than men, aphasic symptoms would be more frequently expected in women than men following right hemisphere lesions. This is not the case (Kimura 1983).

Beyond important questions regarding the reliability of individual reports of sex differences in the brain, another critical issue concerns the relation between structure and function. It is of course critical for Baron-Cohen's thesis that structural sex differences have functional implications for cognition and behavior relating to empathizing and systemizing.[11] However, the functional meaning of the structural differences mentioned by Baron-Cohen and colleagues (2005) remains unclear, and have not been directly associated with empathizing or systemizing skills. Baron-Cohen et al. attempt to relate structural sex differences to differences in function, but their links are hypothetical at best. For example, they tentatively suggest that a male skew toward local rather than long-range connectivity would be a disadvantage for empathizing, "because empathy activates brain regions that integrate information from multiple neural sources" (2005: 821, reference

removed) and that "[t]his notion of skewed connectivity is also compatible with strong systemizing, because systemizing involves a narrow attentional focus to local information, in order to understand each part of a system."

While we would not disagree that empathizing requires integration of information from many regions of the brain, we would dispute that there is any reason why systemizing, or any other complex behavior, would require any less integration. A local focus in the mind does not imply a local focus in the brain. The complexity of the relationship between brain structure and function is immense, and as Fausto-Sterling (2000b: 118; see also Bleier 1986) has observed, "despite the many recent insights of brain research, this organ remains a vast unknown, a perfect medium on which to project, even unwittingly, assumptions about gender."

Overall then, links between purported structural differences and functional differences pertaining to E/S currently remain speculative at best. Hypothesized sex differences in lateralization, corpus callosum size, and proportions of gray and white matter are under empirical dispute, highlighting the importance of treating with skepticism isolated reports of sexual dimorphism in the brain. Moreover, the functional significance of any such differences currently remains unknown.

3.4 Summary of evidence that fT organizes functional brain 'type'

In the preceding sections we have examined three categories of evidence claimed to support the idea that (purported) behavioral sex differences in E/S profile are partially caused by the action of fT on the brain: studies of girls with CAH; studies of aT and E/S profile in the general population; and studies of brain sexual dimorphism. Our analysis shows that in no domain of research do the data provide anything like compelling support for the E/S hypothesis. Importantly, the causal links between fT, brain organization, and cognitive profile, on which the E/S hypothesis hinges, are never demonstrated. No evidence is provided to suggest how fT is responsible for structural sex differences in the brain, or how these differences are responsible for differences in E/S profile. When links are made, they are highly speculative. Moreover, the studies of both fT and E/S, and of brain sexual dimorphism, have provided patterns of results that are highly inconsistent.

4 The E/S hypothesis: summary

The E/S hypothesis attempts to explain sex differences in cognitive style and behavior in terms of fT's organizational effects on the brain. A

careful analysis of the evidence provided by Baron-Cohen et al. (2005) in support of their theory, as well as additional and more recent research, reveals that such evidence is far from being convincing. Purported sex differences are irrelevant to E/S profile, under dispute, are eliminated in certain social contexts, or may be due to experiential factors. The studies of aT in humans, promising at first, have not provided reliable evidence of the role of fT on behavioral and cognitive sex differences. Furthermore, these studies are often tainted by serious methodological flaws, and the relationship between proxies for fT and actual fT exposure is unknown. Sexual dimorphism in the human brain, other than size, is not yet established, and has not been linked to sex differences in behaviors or cognitive styles, or, importantly, to fT. As previously mentioned, no clear picture emerges of what fT is purported to organize in terms of neural structures, cognitive styles, and behavior. It is of concern that Knickmeyer et al. (2008: 995) have recently claimed that "It is difficult to find any cognitive measure which we can be certain is a proxy measure of fT exposure. If we are to use a cognitive task in this way, the focus should be on tasks where multiple different methodologies implicate fT, including studies of females with CAH, males with androgen insensitivity, correlations with digit ratio, and correlations with amniotic testosterone levels (Baron-Cohen et al. 2005)." To our knowledge, no cognitive measure satisfies these constraints. Even performance on the mental rotation task, the most robust known cognitive sex difference, has not been unequivocally linked to fT or its proxies (Malouf et al. 2006; Puts et al. 2008). In this scenario, no support whatsoever seems to exist for the E/S hypothesis. The authors seem to ignore this obvious conclusion, and instead claim that no cognitive measure might be convincingly considered a behavioral proxy for fT. This approach, scientifically unacceptable, potentially sets the stage for making the E/S hypothesis impossible to test.

In addition to the empirical weakness of the support for the E/S hypothesis, we have also argued that it implicitly subscribes to a conception of development that assumes a unidirectional causal pathway from genes to structural brain changes to psychological function (see Gottlieb 1992, for a critique of this view). As a point for future consideration, developmental cognitive neuroscience is beginning to yield a picture of development characterized by a gradual increase in regional specialization and modularization of function (e.g. Johnson et al. 2005; Karmiloff-Smith 2007). In this view, brain organization emerges through development, through the complex interaction of multiple factors (including behavior itself) rather than being simply the result of

maturation processes. This conception of development doesn't preclude the possibility that, during one brief period of gestation, fT can act on the developing brain in a way that has consequences for future function. However, compared with a 'development to' perspective, it is less plausible that fT levels at a single time-point might have direct and measurable effects on complex psychological function many years later. And indeed, to highlight the theme of this chapter, there is no evidence that supports the E/S hypothesis of fT-directed sexual dimorphism of brain structure and function.

Methodological rigor, measured judgment, and caution should be trademarks of scholars researching in potentially sensitive fields, such as the one of sex differences. As Baron-Cohen (2007: 160) has suggested, "the field of sex differences in mind needs to proceed in a fashion that is sensitive ... by cautiously looking at the evidence and being careful not to overstate what can be concluded." In contrast with this avowed sentiment, we note a frequent lack of acknowledgment of the methodological weaknesses or inconsistency of results that limit the conclusions that can be drawn. Furthermore, the several reference errors and the frequent misrepresentation of results reveal an interpretation of the literature that not only is not cautious but often imprecise or inaccurate.

We end by noting recent evidence that accounts of gender difference that emphasize 'biological' causes are associated with increased endorsement of gender stereotypes (Brescoll and LaFrance 2004), increased self-stereotyping (Coleman and Hong 2008), stereotype threat (Dar-Nimrod and Heine 2006; Thoman et al. 2008), as well as increased confidence that society treats women fairly, reification of the gender status quo, and increased tolerance for sex discrimination in the workplace (Morton et al. 2009). Thus, the empirical and conceptual inadequacies of E/S theory, and its presentation, are of significant social, as well as scientific, concern.

Appendix

Empathizing and aT

All the studies that fall into this category were conducted by Baron-Cohen and colleagues, and they drew on a single population of children whose mothers had amniocentesis. The first study aimed to relate aT with eye contact at 12 months of age (Lutchmaya, Baron-Cohen, and Ragatt 2002a). The infant, in the company of both a parent and the experimenter, was given toys to play with, and eye contact frequency with the parent was used as the dependent variable. (While contact

frequency and contact duration were correlated, the former was considered, without an explanation, a more accurate measure than the latter; see 329.) The authors considered eye contact a "marker of social development" (328), but did not explain in what sense, apart from noting that autistic individuals show fewer eye contacts than individuals without autism. However, in this particular experimental situation, in which the infants interacted with a stranger (the experimenter), eye contact with the parent could reflect shyness, fear, or concern. Indeed, given that the experimenter was a stranger, arguably a better measure of social competence would be eye contact with the experimenter. Lutchmaya and colleagues found a higher frequency of eye contact in females than males. They also found no relationship between aT and eye contact frequency in females. This result was explained by the authors in terms of small sample size (n = 30), but their data, as shown in the scatterplot, shows no relationship between the two measures. In males (n = 41), the function was quadratic, which means that a high frequency of eye contact was observed in males with low and high levels of aT. This result runs contrary to the E/S hypothesis, according to which higher levels of fT should be associated with low frequency of eye contact (as in autism). A number of methodological concerns can also be raised, especially the apparent lack of any attempt to either control or monitor the gaze behavior of either parents or the experimenter. There was also no information regarding whether the experimenter was blind to either the experimental hypothesis, or the infant's aT status. Moreover, each infant was filmed for "approximately" 20 minutes (328); as a result, the frequency of eye contact could have been overestimated in some infants or underestimated in others due to an apparently variable length of observation time.

The second study is an investigation of vocabulary size in 18- and 24-month-old infants (Lutchmaya, Baron-Cohen, and Ragatt 2002b).[12] However, as noted earlier, it is not clear that vocabulary size reflects empathizing ability. Based on mothers' report, females had a larger vocabulary size than males. An inverse relationship was found between aT and vocabulary size when boys and girls were pooled together but not within each sex. Due to small sample size, the authors did not carry out a regression analysis excluding sex.

A later study by Knickmeyer et al. (2005a) tested for relationships between aT and two subscales of the Children's Communication Checklist (CCC, Bishop 1998). The subscale most relevant to empathizing assessed 'quality of social relationships' (tapped by questions such as "is popular with other children"). There was a trend for females to score better on this subscale, as reported by mothers, but it failed to reach significance (however, the effect size was moderate). While a negative relationship was found between these scores and aT with girls and boys pooled together, no relationship was found when the analysis was run within sexes. It is noteworthy that no sex differences were found on the pragmatic subscale of the CCC (which measures how children adapt to their interlocutors during a conversation), although the authors predicted "that higher fT levels would be associated with poorer scores on the pragmatic language scale" (Knickmeyer et al. 2005a: 200).

Baron-Cohen and colleagues (Chapman et al. 2006) also explored the relationship between aT and a children's version of the Empathy Quotient (EQ-C, filled out by mothers), comprising questions such as "My child shows concern when others are upset" and "My child can easily tell when another person wants to enter into conversation with him/her." Girls were rated as higher in empathizing skills, but maternally reported EQ-C score was not validated against any social performance measure. Correlations between aT and EQ-C revealed a negative relationship between aT and EQ-C for girls and boys combined. When the two groups were examined separately, a significant negative correlation was found only in boys. Chapman et al. (2006) also made use of a performance measure, a child's version of the "Reading the Mind in the Eyes" test (Eyes-C). In this test the child is shown just the eye region of a series of faces, and is offered four multiple choice options as to what the individual is thinking or feeling. Interestingly, females did not perform better than males on the task, and the authors note that they previously failed to find female superiority on the task (Chapman et al. 2006: see 140). Analyzes revealed a significant negative correlation between aT and Eyes-C score, in both boys and girls, and in both sexes separately. However, in the absence of a sex difference in behavior, it is not clear that these findings as a whole can be taken as support for the E/S hypothesis.

Knickmeyer et al. (2006b) also used a performance measure, involving two computer-presented films in which animated shapes move in ways that convey that they have mental states (see Abell, Happé, and Frith 2000). Children were probed by an interviewer to describe the events of the animation, and prompted to do so in terms of human interactions. No information is provided as to whether or how this probing and prompting was standardized for all children (see Knickmeyer et al. 2006b: 285 for sample transcript of interview), and there is no information regarding whether the interviewer was blind either to the experimental hypothesis or aT status. The authors predicted that females

would use more mental and affective state terms than boys, as well as more reference to actions between animate objects (e.g. "The big one's trying to hit the little one"), and more intentional propositions (which included mental and affective state terms referring to emotional states, beliefs, and desires). These predictions were partially supported: significantly greater affective state term use was seen in females, and there was a trend for greater female use of intentional propositions. However, females did not use mental state terms more, or make greater reference to actions between animate objects, and there was an unpredicted difference in the use of neutral propositions (which was greater in boys). The authors then went on to investigate the relationships between aT and these four dependent variables.

Analyzes revealed no association between aT and mental and affective state terms, and no correlation within either sex. For both intentional and neutral propositions, main effects of aT were seen, with higher aT being associated with less intentional proposition use, and greater neutral proposition use. For intentional propositions, a negative correlation between the two variables was seen within boys, but not girls. For neutral propositions, no correlations were seen within either sex. Summarizing their findings, Knickmeyer et al. (2006b: 288) state that they "predicted that females would use more mental and affective state terms than males" as well as more intentional propositions, and that "variation in fT levels would account for the predicted sex differences. In general, our predictions were supported." We would argue that this is an overly optimistic conclusion. Of their four predicted sex differences on the task (overlooked in their summary is their prediction that females would refer more to actions between animate objects), only one attained statistical significance (affective state term use). This variable showed no correlation with aT levels. A relationship between intentional propositions (only marginally more frequent in females) and aT was established but, within sex, a correlation was seen only in males. We would suggest that no firm conclusions can be drawn from this study.

Systemizing and aT

We turn now to the smaller number of studies that have explored relations between aT and purported measures of systemizing. The earliest is a study of four-year-old children who were assessed on a range of cognitive tasks, and scores related to aT (Finegan, Niccols, and Sitarenios 1992). In girls, where relationships with aT were observed, they were contrary to the predictions of the E/S hypothesis: relationship with classification abilities was curvilinear; higher aT was associated with lower score on counting, number facts, and block building scores. No

relationships with abilities such as puzzle solving, visual-motor integration, and embedded figures were found. In boys, no relationships between aT and any cognitive abilities were observed.

A later study tested seven-year-old girls and boys on a mental rotation task study (Grimshaw et al. 1995). As noted earlier (see section 2.1), it is unclear whether mental rotation should be understood as a measure of systemizing. The sample was small and a correlation between levels of aT and performance speed was found only in girls, and only for those girls who employed a rotation strategy (n = 12). Importantly, there were no overall sex differences in accuracy or RT (although among rotators girls were faster, whereas boys were faster among non-rotators). As noted by Hines (2007), it is performance accuracy – that did not relate to aT – on which a sex difference is normally seen. This article therefore does not provide convincing evidence of a relationship between aT and mental rotation. More recently, Auyeung et al. (2009a) tested for a relationship between aT and block design performance (thought to assess visuospatial skill). Contrary to prediction, no male advantage was observed for block design performance, nor any relationship with aT.

One of the subscales from the Children's Communication Checklist (Bishop 1998) used by Knickmeyer et al. (2005a) tested restricted interests, considered associated with systemizing abilities and autism according to the E/S hypothesis. The scale inquired about the presence of specific interests (e.g. "has one or more over-riding specific interests (e.g. computers, dinosaurs) and will prefer doing activities involving this to anything else"; Bishop 1998: 891) and, once again, was filled out by mothers. The authors predicted that boys would show more restricted interests than girls and that more restricted interests would be associated with higher levels of aT. A sex difference in the predicted direction was found. A relationship between restricted interests scores and aT was found when the two sexes were pooled together; in within-sex analysis, a relationship was found in boys but not in girls. At page 205, the authors stated "There was a main effect of fT on this scale when the group was examined as a whole... This indicates that in both boys and girls, higher fT levels are associated with more restricted interests." This statement is inconsistent with the results; indeed, although the regression analysis for the pooled sexes indicated a positive relationship between the two variables, the correlation was not significant for girls. Moreover, its sign was negative.

Another issue that can be raised concerns the nature of the scale items. The content of some items is clearly sex-related. For example, one

capitals of the world, names of many varieties of dinosaurs), or overriding specific interests (e.g. computer or dinosaurs). It is not clear why restricted interests for social interactions are not included, or why knowing all the names of the children in kindergarten is not an example of restricted interest. Furthermore, it is not clear why "prefers to be with adults rather than other children" is associated with restricted interests. In a nutshell, it seems that some of these items are associated with what are considered sex-typical behaviors and not restricted interests *per se*.

One final study exploring the relationship between aT and systemizing made use of a version of the Systemizing Quotient, adapted for parents to rate their children (the SQ-Child, Auyeung et al. 2006). Parents of boys gave higher ratings on the SQ-Child, on average, than parents of girls. Moreover, aT was significantly associated with SQ-Child score, and within-sex correlations were significant in both boys and girls. While this appears to provide strong support for the E/S hypothesis, inspection of the items of the SQ-Child (see Auyeung et al. 2006: S126) raises the serious question of whether it actually taps systemizing ability. In addition to the subjectivity of parental report as opposed to actual performance, very few of the items are unambiguous tests of "the drive to analyze or construct systems" (Auyeung et al. 2006: S124). At least half of the 28 items appear to tap into a drive for order, routine, or arrangement of objects. Nor is it clear in many (if not all) of the remaining items that it is a drive to systemize that is being tapped (e.g. items include knowing the difference between the latest models of game-consoles, finding using a computer difficult, enjoying working to solve a puzzle, or spending time mastering aspects of their favorite activities).

Gender-typical play

To date four articles have investigated the relationship between measures of fT and gender-typical play behavior. As with the studies with girls with CAH, discussed in section 3.2.1, no attempt has been made in this research to test a specific hypothesis regarding what it is about toys and play behavior culturally ascribed to boys versus girls that makes them differentially appealing to a more or less 'masculinized' brain. Again, too, play behaviors that fail to elicit the predicted difference between the sexes may simply be replaced (Knickmeyer et al. 2005b). The first study assessed behavior using the Pre-School Activities Inventory (PSAI, Golombok and Rust 1993), and used both maternal T and maternal sex hormone-binding globulin (SHBG) levels as potential proxies for fT

functional effectiveness by binding with it, greater levels of SHBG is used as a proxy for lower levels of unbound, functionally effective T.) Preliminary analyzes confirmed that, as noted earlier, levels of maternal T and maternal SHBG did not differ in mothers bearing male versus female fetuses, underlining its questionability as an adequate proxy for fetal exposure. Hines et al. (2002) found that, in girls only, higher levels of maternal T (but not maternal SHBG) were associated with more masculine-typical play. No other relationships were significant.

Subsequently, Knickmeyer et al. (2005b) looked for a relationship between aT and gender-typical play in four- and five-year-old children, as measured by a questionnaire about play behavior, filled out by the mothers. No relationship with aT was found in either sex, or in both sexes together. Van de Beek et al. (2009) explored relationships between both maternal and amniotic T, estradiol, and progesterone levels and actual gender-typical play in 13-month-old infants in the laboratory. They found no relationship with amniotic or maternal T or estradiol levels. Surprisingly, higher levels of amniotic progesterone were associated with a stronger preference for masculine toys. However, in contrast with the largely negative findings of these three studies Auyeung et al. (2009b), with a larger sample size, found correlations, in both sexes individually as well as pooled, between aT and PSAI score. It will be important to establish whether this result can be replicated and to investigate whether it is an intrinsic difference between 'masculine' and 'feminine' activities that is correlated with aT, or their social ascription to gender. (It is also worth noting that if aT levels are influenced by maternal T levels, then a correlation between aT and play preferences may be mediated by differential social experiences provided by mothers who are lower or higher in T.)

Notes

The authors thank Alison Nash and Rebecca Jordan-Young for their very valuable comments on an earlier draft of this chapter. The authors contributed equally to this chapter.

1. Note that this article does not refer to sex differences in mental rotation ability.
2. This citation is unclear. Baron-Cohen et al. (2005) referred to the Witkin book *Personality Through Perception* in their References list. The book was published in 1972, not 1962, and had Witkin as the only author. It is likely that the authors were referring to Witkin, Dyk, Faterson, Goodenough, and Karp (1962/1974), *Differentiation: Studies of Development* (Hoboken, NJ: John Wiley). Because of this imprecision, we have not included this citation in