Brief Communications

Cooperation and Heterogeneity of the Autistic Mind

Wako Yoshida,[1] Isabel Dziobek,[2] Dorit Kliemann,[2,3,4] Hauke R. Heekeren,[2,3,4] Karl J. Friston,[1] and Ray J. Dolan[1]
[1]The Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London WC1N 3BG, United Kingdom, [2]Cluster of Excellence "Language and Emotion" and [3]Department of Educational Science and Psychology, Freie Universität Berlin, 14195 Berlin, Germany, and [4]Max Planck Institute for Human Development, 14195 Berlin, Germany

Individuals with autism spectrum conditions (ASCs) have a core difficulty in recursively inferring the intentions of others. The precise cognitive dysfunctions that determine the heterogeneity at the heart of this spectrum, however, remains unclear. Furthermore, it remains possible that impairment in social interaction is not a fundamental deficit but a reflection of deficits in distinct cognitive processes. To better understand heterogeneity within ASCs, we employed a game-theoretic approach to characterize unobservable computational processes implicit in social interactions. Using a social hunting game with autistic adults, we found that a selective difficulty representing the level of strategic sophistication of others, namely inferring others' mindreading strategy, specifically predicts symptom severity. In contrast, a reduced ability in iterative planning was predicted by overall intellectual level. Our findings provide the first quantitative approach that can reveal the underlying computational dysfunctions that generate the autistic "spectrum."

Introduction

Despite recent progress in social and affective neuroscience, there are no simple models that explain and predict complex human social behavior and its pathologies. Autism spectrum conditions (ASCs) involve severe deficits in reciprocal social interaction (American Psychiatric Association, 1994) and thus represent a behavioral deficit model for social functioning. A lack of "theory of mind" (ToM), i.e., the representation of others' goals and intentions (Frith and Happé, 2005), is thought to be a key pathognomonic feature defining social dysfunctions of ASCs. Although many tasks are sensitive to impairments in various facets of ToM (Tager-Flusberg et al., 2001), a systematic approach to the common underlying mechanism has yet to be realized. Furthermore, it is widely assumed that aberrant ToM is not a fundamental deficit in itself but rather a reflection of multiple underlying cognitive processes, each of which may be manifested in different forms, or subtypes, of ASCs. Here, we employed a game-theoretic approach which has recently provided new insights into the computations and neural correlates of social behaviors (King-Casas et al., 2008; Behrens et al., 2009).

We used a Stag-hunt game in which participants interacted with a computerized agent to hunt stags together (high value) or defect to hunt rabbits alone (low value) (Yoshida et al., 2008) (Fig. 1). In the Stag-hunt, cooperation depends on recursive representations of another's intentions, since, if I decide to hunt the stag, I must believe that you believe that I will cooperate with you; thus interactions of highly sophisticated players allow cooperation to emerge (Yoshida et al., 2008). During the experiment, a computerized agent shifted its sophistication (by three degrees of recursion) without notice. Thus, to behave optimally, participants were required to (1) estimate the agent's sophistication level, (2) update their own strategies continuously, and (3) behave optimally on the basis of their inference. This calls on recursive belief inference, cognitive flexibility, and interactive planning, respectively, all of which have been reported to be impaired in ASCs (Sigman et al., 2006). To tease apart tightly entangled cognitive processes implicit in social interaction, we used a previously developed theoretical model, in which participants behave optimally with respect to the goal of maximizing the payoff based on these three processes. This model-based behavioral analysis allowed us to capture the linkage of unobservable computational processes with observable intellectual and diagnosis scores that contain the heterogeneity within the autism spectrum.

Materials and Methods

Participants. Seventeen adults with ASCs (12 males) and 17 healthy control adults (11 males) participated in the study. All participants were German speaking, and age (ASC, 33.1 ± 7.8 years; control, 31.0 ± 5.7 years) was matched between the two groups. For each participant in the ASC group, diagnoses were made according to DSM-IV criteria (American Psychiatric Association, 1994) using a videotaped semi-structured interview. All participants gave informed written consent and the research protocol was approved by the Max Planck Institute for Human Development, Berlin, Germany.

Diagnostic and intellectual measures. To compare the ASC and control subjects in autistic diagnosis, we used the autism-spectrum quotient (ASQ), a screening questionnaire used in clinical practice. The ASQ score of the control group (mean ± SD $= 10.8 \pm 5.2$) was significantly lower than that of the ASC group (38.7 ± 8.0, $p < 0.1 \times 10^{-12}$) (Fig. 2*A*). We also confirmed that all control individuals scored lower than 26 and were ruled out as ASC. For the participants in the ASC group, autistic symptomatology was quantified using the Autism Diagnostic Interview-Revised (ADI-R) (Lord et al., 1994) in 16 individuals and the Asperger Syndrome and High Functioning Autism Diagnostic Interview (ASDI) (Gillberg et al., 2001) in 14 individuals. The ADI-R is a semi-structured interview administered to the parents or the caretakers, while the ASDI is designed to cover diagnostic criteria for Asperger and is performed by the individuals with autism themselves.

Received Jan. 24, 2010; revised March 31, 2010; accepted April 26, 2010.

This work was supported by Wellcome Trust Programme grants to R.J.D. and K.J.F.

Correspondence should be addressed to Wako Yoshida, The Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL, 12 Queen Square, London WC1N 3BG, UK. E-mail: w.yoshida@fil.ion.ucl.ac.uk.

DOI:10.1523/JNEUROSCI.0400-10.2010

The intelligence level of the participants in the ASC and the control group was measured by a vocabulary test [Mehrfachwahl-Wortschatz-Test or MWT (Tewes, 1991)] and a nonverbal strategic thinking test [Leistungsprüfsystem or LPS (Horn, 1962)]. There was no significant difference between the two groups for both verbal (control, 117.1 ± 15.8; ASC, 105.9 ± 16.0) and nonverbal (strategic) intelligence scores (control, 124.2 ± 13.3; ASC, 124.2 ± 12.0) (Fig. 2*A*). To assess fluid and crystallized intellectual functioning of the ASC, we built the mean of the two scores, and based on these, WAIS-R full-scale intelligence quotient (IQ) was estimated.

Experimental task and computational model. We used a Stag-hunt game (Skyrms, 2003) in which participants interacted with a computerized agent, thereby using a ToM model, which is a generative model to infer participants' mental states in the Stag-hunt. As the game and the model have been described in detail previously (Yoshida et al., 2008), we briefly summarize the key concepts here.

The Stag-hunt game has a pro-cooperative payoff matrix, in which each of two players choose whether to hunt highly valued stags together (20 points each) and share the proceeds, or defect to hunt rabbits of smaller value (10 points). The task was embedded into a move-by-move maze-based hunting game (Fig. 1), where each game finishes when either player has caught a prey. Each move incurs a small cost (1 point), such that the best strategy is to hunt bigger prey as swiftly as possible. In the experiment, the participants played two sessions, each of which comprised 40 games. At the end of the experiment, the participants were paid money based on accrued total points.

The player's strategy is defined by the hierarchical value-functions using optimal control theory. In sequential games with multiple players, like the Stag-hunt game, the payoffs are defined over a joint-space for each player and the values for one player become a function of the other player's values. Thus the optimal value-functions (strategies) are specified by an order based on recursive sophistication. Namely, the first-order strategies discount the strategies of other players (i.e., I will ignore your goals). Second-order strategies are optimized under the assumption that other players are using a first-order strategy (i.e., you are ignoring my goals). Third-order strategies pertain when I assume that you assume I am using a first-order strategy and so on. In general, a player with the lower-order (competitive) strategy would try to catch a rabbit, provided both players were not close to the stag, and a player with a higher-order (cooperative) strategy would tend to chase the stag even if it was close to a rabbit. In the experiment, the computer agent changed strategy, which is defined with first-, third-, or fifth-order sophistication, without notice.

Using the hierarchical value-functions, we tested two computational models of a player's behavior. The ToM model assumes that players infer the other's strategy; the sophistication level of the computerized agent in the Stag-hunt, based on Bayesian inference under flat priors bounded by K (bounded rationality) after each observation (moves in the maze at each trial):

$$\text{agent's level}(T) = \underset{\text{agent's level} \in \{1:K\}}{\arg\max} \prod_{t=1}^{T-1} \kappa^{T-1} \times p(\text{observation}(t+1) \mid \text{subject's level}(t), \text{agent's level}(t)),$$

where κ is a forgetting parameter that exponentially discounts previous evidence and allows the player to respond more quickly to changes in the other's strategy. The parameter value ranges from zero, the most flexible setting in which players update the strategy based only on the latest action of others, to one, when players take the history of all actions into account. Accordingly, under the definition of value-functions with order, the players optimize their own strategy, which is one level higher than the estimated other's sophistication level. As a reference model without belief inference, we assumed a "fixed strategy" model in which the players use a fixed strategy at the level of k and do not change strategy based on others' behavior.

Model selection and parameter estimation. We computed the evidence of the fixed strategy models, $p(M_k^{\text{FIX}})$, with $k = 1, \ldots, 6$ and the ToM models, $p(M_K^{\text{TOM}})$, with $K = 1, \ldots, 6$; i.e., 12 models in total, given the actual behavioral data. To evaluate the expectation of belief inference, we

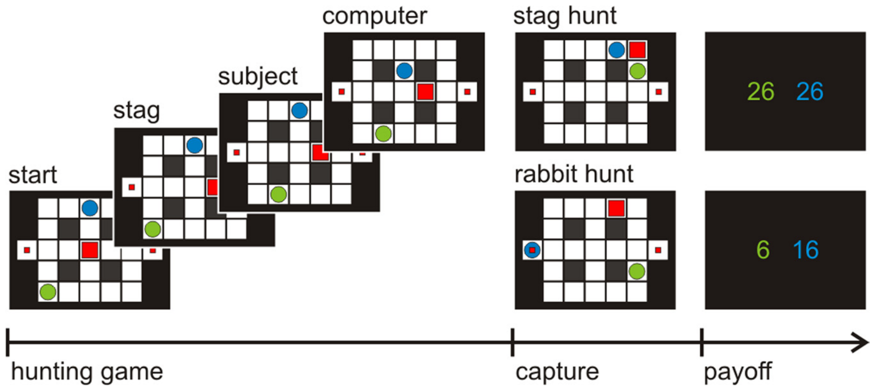

Figure 1. Stag-hunt game. Two players, a subject (green circle) and a computer agent (blue circle), try to catch prey: a mobile stag (big square, big payoff) by cooperation or two stationary rabbits (small squares, small payoff), by moving in a sequential manner. At the end of each game, both players receive points equal to the sum of prey and points relating to the remaining time.

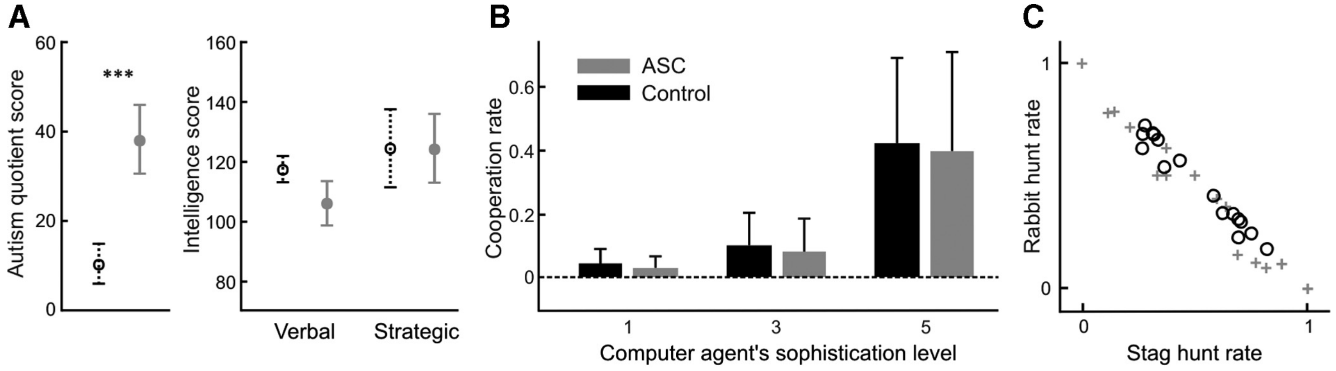

Figure 2. Subjects and behavioral results. ***A***, Mean scores of diagnostic (left panel) and intellectual measurements (right panel) of the control group (black dotted line) and the ASC group (gray solid line). The error bars show the SDs. ASQ score was significantly higher for the ASC group than for the control group ($p < 0.1 \times 10^{-12}$), while there was no significant difference between the groups for both verbal and strategic intelligence scores. ***B***, Both the control and the ASC group attempted to catch a stag, in effect, when they behaved more cooperatively, when the computer agent was more sophisticated. ***C***, The rate of stag hunt and rabbit hunt in the games with sophisticated computer agent with the fifth-order strategy. The participants in the ASC group (gray crosses) showed a larger variety of behavior than the control participants (black circles).

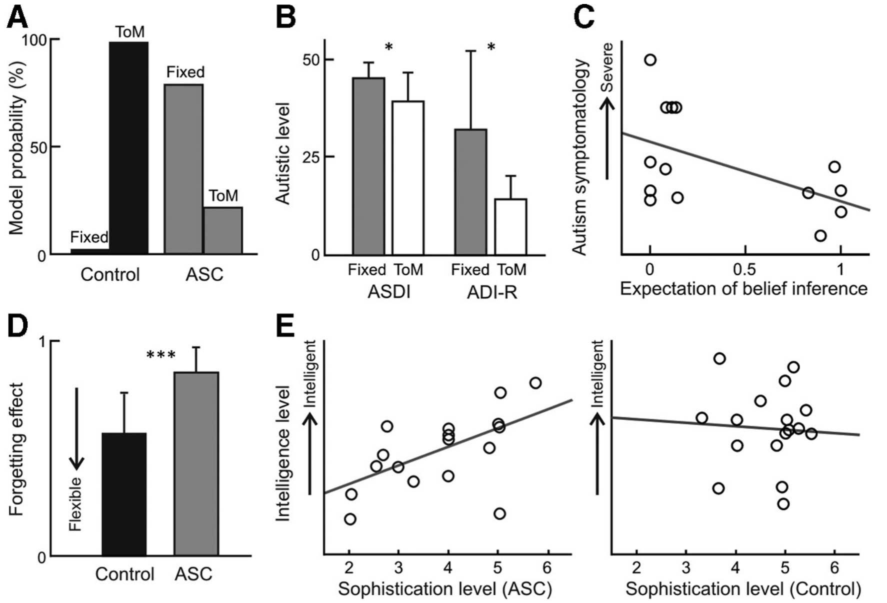

Figure 3. Model-based behavioral results. ***A***, The probability for the ToM model (98.2%) was higher than that for the fixed strategy model for controls, while the fixed strategy model was dominant (78.6%) in individuals with ASCs. ***B***, Diagnosis measurement scores, ADI-R and ASDI, were significantly higher for the ASC participants whose behavior fit better with the fixed strategy model ($n = 12$) than those showing a better fit with the ToM model ($n = 5$). ***C***, In the ASC group, the greater the expectation of recursive belief inference, the more severe was the autism symptomatology ($n = 14, r = -0.52, p = 0.055$), as measured by the sum of scores on the ADI-R and the ASDI. ***D***, The estimated forgetting parameter for the ASC group (mean ± SD = 0.57 ± 0.19) was significantly higher than that for the control group (0.93 ± 0.13) ($p < 0.1 \times 10^{-6}$). ***E***, The estimated sophistication for the individuals with autism showed significant positive correlation with individual IQ scores (left panel: $n = 17, r = 0.54, p = 0.026$), while there was no correlation for the control participants (right panel: $n = 17, r = 0.02$).

calculated the marginalized posterior probabilities of the fixed strategy model and the ToM model by accumulating evidence for different models specified by different k-level and bound K; i.e., 6 models each, as $\sum_k p(M_k^{\mathrm{FIX}})$ and $\sum_k p(M_k^{\mathrm{TOM}})$.

To estimate the cognitive flexibility, we used the average forgetting effect, which is calculated as the weighted sum of the parameter κ with the posterior probability of each model. The forgetting parameter was optimized for each ToM model with different K, and set at one for the fixed strategy models:

$$\sum_k p(M_k^{\mathrm{FIX}}) + \sum_k \kappa_K p(M_k^{\mathrm{TOM}}).$$

The average sophistication for the ASC participants was estimated as the weighted sum of parameter k with the model probability of each fixed strategy model.

Results

To first confirm a basic cognitive understanding of the task demands, we examined behavioral strategies in the ASC group. Overall behavioral measurements including the average reaction times, total earnings, and cooperation rates of the ASC group did not differ from the control group, providing evidence that an understanding of the social task was intact (Chiu et al., 2008; Uddin et al., 2008). The participants in the control group attempted to catch a stag, i.e., they behaved more cooperatively, when the computer agent was more sophisticated, and this tendency was also observed for the ASC group on average (Fig. 2*B*). The behavioral profiles, however, differed greatly in individuals with ASCs. Thus, while all control participants dealt with both hunting a stag and hunting a rabbit according to the situation (e.g., initial positions), there are extreme participants in the ASC group who never cooperate or compete (Fig. 2*C*). This suggests a behavioral diversity or even multiple phenotypes among the ASC participants.

To identify functional abnormalities in the computational processes involved in the task, we used the ToM model and the fixed strategy model. The ToM model includes two model parameters characterizing the cognitive processing: one is the upper bound of sophistication (K), which defines the capacity of strategic planning, and the other is a forgetting effect, which controls how quickly a player responds to changes in the other's sophistication, thereby representing a measure of cognitive flexibility. For the fixed strategy model, as it is assumed that players do not change their strategy, only the sophistication level (k) is estimated.

First, to compare the behavioral fit of these two models, we calculated the log likelihoods of the models for the control and the ASC participants (Control: ToM −5776, fixed −6060; ASC: ToM −5330, fixed −4937; greater values indicate better fit). Bayesian model selection based on these log likelihoods showed that, for the control participants, the ToM model with belief inference accounted for the behavior significantly better than the fixed strategy model without belief inference. Conversely, the fixed strategy model explained individual behavior better for 12 of 17 participants with ASCs (Fig. 3*A*). We also evaluated the quality of model fitting with a pseudo-r^2 statistic, defined as $(m_0 - m)/m_0$, where m and m_0 are the log likelihood of the data under the model and under purely random choices, respectively. The results showed that the ToM model ($r^2 = 0.230$) provided a better fit than the fixed strategy model (0.194) for the controls, while the fixed model (0.262) provided a better fit than the ToM model (0.203) for the ASC participants.

We divided the ASC participants into two subgroups based on the better explained models, and compared their diagnosis and intellectual scores (Fig. 3*B*). The diagnosis scores (the ADI-R and ASDI) were significantly higher for the ASC participants whose behavior fit better with the fixed strategy model ($n = 12$) than others with the ToM model ($n = 5$), while there was no difference on the intellectual (IQ) scores. Furthermore, the probability of the ToM model, how likely recursive belief inference is used, correlated negatively with individual autistic symptomatology as measured by the sum of scores on the ADI-R and the ASDI (Fig. 3*C*). The correlation was not significant ($p = 0.055$) with the rather small number of sample size ($n = 14$); however, this negative correlation was also observed with the ADI-R ($n = 14, r = -0.45, p = 0.105$) and the ASDI score ($n = 16, r = -0.44, p = 0.090$) individually. We thereby characterized unobservable computational processes implicitly involved in ToM quantifying the individual ability for recursive belief inference.

In terms of the model parameters, the mean estimated forgetting effect in the ASC group was significantly higher than that of the control group (Fig. 3*D*). The higher value of the forgetting effect means that participants are tied to their past strategies, and this result indicates that the ASC participants had an additional impairment in cognitive flexibility. For the sophistication level, we inferred that the upper bound under the ToM model (K) is equal to 5 for 13 of 17 control participants, which means the participants used strategies with up to a sixth sophistication level. This is reasonable given that the computer agent's policies never exceeded level 5. In contrast, we found that the average sophistication under the fixed model (k) varied among the ASC participants (mean ± SD = 4.3 ± 1.2) and correlated positively with individual IQ scores (Fig. 3*E*). This correlation was not observed for control participants' IQ scores and the sophistication level, which is calculated as the expected upper bound of the ToM models (Fig. 3*E*). Note that, from our model-based behavioral analysis, we concluded that the ASC participants have a difficulty in changing their sophistication level (k-level), while the controls flexibly change it, but under the fixed upper bound (K). Thus, these parameters of sophistication levels (K and k) are functionally different for the two groups. However, we obtained almost identical results for both the ASC and control group using the expected sophistication and its bound using Bayesian model averaging over both fixed and ToM models.

Discussion

In summary, our model-based approach isolates specific factors that contribute to ASC social impairments, which parametrically covary with symptom severity and intelligence level. Thereby our findings offer a partial account of the heterogeneity of autism spectrum conditions.

While we found evidence that the general understanding of our social Stag-hunt task was intact in ASCs, the observed game behavior of individuals with ASCs was guided to a significantly lower degree of belief inference than that of the control participants. Moreover, severity of autism symptomatology predicted the extent to which ASC participants behaved according to a fixed strategy, i.e., disregarded the other player's beliefs. This is line with research showing that ToM impairments in autism are core deficits and underlie the individual's social dysfunction (Tager-Flusberg et al., 2001; Baron-Cohen and Belmonte, 2005).

Individuals with autism also showed deficits in cognitive flexibility as indicated by a tendency to be tied to their past strategies during the social game rather than a policy of flexibly changing strategies by taking into account the other's new actions. In addition, the level of sophistication (you think that I think that you think, etc.) for participants in the ASC group was related to IQ scores. A previous study using a "Beauty Contest" game has indicated an association between higher-level reasoning in healthy individuals and higher intelligence scores (Coricelli and Nagel, 2009). Our result showed that highly intelligent individuals with ASCs behave cooperatively as if they make predictions over a longer time-horizon. This suggests that the level of sophistication, a key component of higher-level reasoning, can be inferred in more complex dynamic social exchanges.

Impairments in cognitive flexibility and planning are known characteristics of neuropsychological profiles in ASCs and have fostered the development of theories of executive dysfunction in autism (Hill, 2004; Russo et al., 2007). However, the precise contribution and nature of the relationship between ToM and executive functions remain a matter of debate (Beauchamp and Anderson, 2010). Using our Stag-hunt game, we show that the domains of both belief inference and executive function contribute to determining social function. Future studies with our model-based approach could potentially disentagle ToM from executive functions and shed light on the question of how these cognitive dysfunctions influence one another and generate autism social symptomatology.

Recent studies using game-theoretic frameworks have shown that frontal brain regions such as the medial prefrontal cortex, orbitofrontal cortex, and cingulate gyrus, as well as the amygdala and superior temporal sulcus are engaged by social exchanges (Sanfey, 2007; Hampton et al., 2008; Lee, 2008). Function in these regions has been repeatedly reported as abnormal in ASCs (Frith, 2001; Amaral et al., 2008). However, none of the studies to date throws light on how different subtypes or symptomatic manifestations in ASCs are determined by dysfunction in distinct regions, and how abnormal connectivity between these regions results in deficient social reciprocal behavior. Our findings highlight the power of simple assessment tools for psychopathology predicated on the idea that core psychiatric deficits can be understood in terms of computational dysfunction. Thus, future studies might fruitfully explore how the specific mechanistic deficit that we observe here relates to function or dysfunction in the social brain.

References

Amaral DG, Schumann CM, Nordahl CW (2008) Neuroanatomy of autism. Trends Neurosci 31:137–145.

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, Ed 4. Washington DC: American Psychiatric Association.

Baron-Cohen S, Belmonte MK (2005) Autism: a window onto the development of the social and the analytic brain. Annu Rev Neurosci 28:109–126.

Beauchamp MH, Anderson V (2010) SOCIAL: an integrative framework for the development of social skills. Psychol Bull 136:39–64.

Behrens TE, Hunt LT, Rushworth MF (2009) The computation of social behavior. Science 324:1160–1164.

Chiu PH, Kayali MA, Kishida KT, Tomlin D, Klinger LG, Klinger MR, Montague PR (2008) Self responses along cingulate cortex reveal quantitative neural phenotype for high-functioning autism. Neuron 57:463–473.

Coricelli G, Nagel R (2009) Neural correlates of depth of strategic reasoning in medial prefrontal cortex. Proc Natl Acad Sci U S A 106:9163–9168.

Frith U (2001) Mind blindness and the brain in autism. Neuron 32:969–979.

Frith U, Happé F (2005) Autism spectrum disorder. Curr Biol 15:R786–R790.

Gillberg C, Gillberg C, Råstam M, Wentz E (2001) The Asperger Syndrome (and high-functioning autism) Diagnostic Interview (ASDI): a preliminary study of a new structured clinical interview. Autism 5:57–66.

Hampton AN, Bossaerts P, O'Doherty JP (2008) Neural correlates of mentalizing-related computations during strategic interactions in humans. Proc Natl Acad Sci U S A 105:6741–6746.

Hill EL (2004) Executive dysfunction in autism. Trends Cogn Sci 8:26–32.

Horn W (1962) Leistungsprüfsystem (LPS) (Achievement Testing System). Göttingen, Germany: Hogrefe.

King-Casas B, Sharp C, Lomax-Bream L, Lohrenz T, Fonagy P, Montague PR (2008) The rupture and repair of cooperation in borderline personality disorder. Science 321:806–810.

Lee D (2008) Game theory and neural basis of social decision making. Nat Neurosci 11:404–409.

Lord C, Rutter M, Le Couteur A (1994) Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord 24:659–685.

Russo N, Flanagan T, Iarocci G, Berringer D, Zelazo PD, Burack JA (2007) Deconstructing executive deficits among persons with autism: implications for cognitive neuroscience. Brain Cogn 65:77–86.

Sanfey AG (2007) Social decision-making: insights from game theory and neuroscience. Science 318:598–602.

Sigman M, Spence SJ, Wang AT (2006) Autism from developmental and neuropsychological perspectives. Annu Rev Clin Psychol 2:327–355.

Skyrms B (2003) The stag hunt and the evolution of social structure. Cambridge: Cambridge UP .

Tager-Flusberg H, Joseph R, Folstein S (2001) Current directions in research on autism. Ment Retard Dev Disabil Res Rev 7:21–29.

Tewes U (1991) Hamburg-Wechsler-Intelligenztest für Erwachsene, Revision. Bern, Switzerland: Huber.

Uddin LQ, Davies MS, Scott AA, Zaidel E, Bookheimer SY, Iacoboni M, Dapretto M (2008) Neural basis of self and other representation in autism: an FMRI study of self-face recognition. PLoS One 3:e3526.

Yoshida W, Dolan RJ, Friston KJ (2008) Game theory of mind. PLoS Comput Biol 4:e1000254.