Biological Psychiatry: CNNI

Archival Report

Neurocognitive Mechanisms of Social Inferences in Typical and Autistic Adolescents

Gabriela Rosenblau, Christoph W. Korn, Abigail Dutton, Daeyeol Lee, and Kevin A. Pelphrey

ABSTRACT

BACKGROUND: Many of our efforts in social interactions are dedicated to learning about others. Adolescents with autism have core deficits in social learning, but a mechanistic understanding of these deficits and how they relate to neural development is lacking. The present study aimed to specify how adolescents with and without autism represent and acquire social knowledge and how these processes are implemented in neural activity.

METHODS: Typically developing adolescents (n = 26) and adolescents with autism spectrum disorder (ASD) (n = 20) rated in the magnetic resonance scanner how much 3 peers liked a variety of items and received trial-by-trial feedback about the peers' actual preference ratings. In a separate study, we established the preferences of a new sample of adolescents (N = 99), used to examine population preference structures. Using computational models, we tested whether participants in the magnetic resonance study relied on preference structures during learning and how model predictions were implemented in brain activity.

RESULTS: Typically developing adolescents relied on average population preferences and prediction error updating. Importantly, prediction error updating was scaled by the similarity between items. In contrast, preferences of adolescents with ASD were best described by a No-Learning model that relied only on the participant's own preferences for each item. Model predictions were encoded in neural activity. Typically developing adolescents encoded prediction errors in the putamen, and adolescents with ASD showed greater encoding of own preferences in the angular gyrus.

CONCLUSIONS: We specified how adolescents represent and update social knowledge during learning. Our findings indicate that adolescents with ASD rely only on their own preferences when making social inferences.

https://doi.org/10.1016/j.bpsc.2020.07.002

Social competences become increasingly important in adolescence—a transition period in which peers begin to outweigh family influence (1). Despite the importance of peer socialization, adolescents make suboptimal social inferences, and in part this is likely due to ongoing development of brain regions (2–4). Specifically, subcortical regions involved in emotion and reward processing, such as the striatum, mature earlier than prefrontal regions, which support cognitive control and emotion regulation (5,6). The medial prefrontal cortex (MPFC), a key region for learning from social feedback, is among the latest maturing brain regions (7,8). In adolescents, the MPFC represents social predictions to a lesser degree than in adults, and that may account for a slower updating of social predictions through environmental feedback (9,10).

Adolescents with autism spectrum disorder (ASD) exhibit difficulties with social cognition (11,12). Social attention and motivation accounts hold that reduced orientation to social stimuli early on in infants with ASD leads to cascading effects on social learning and social interaction later on (13–15). With respect to adolescents with ASD, it is not clear whether social attention and motivation continue to change in this critical period for brain development.

Reduced attention and motivation to social stimuli would certainly lead to an impoverished representation of social knowledge. Theories about knowledge acquisition postulate that humans acquire and represent abstract knowledge automatically by extracting statistical regularities from situations based on a set of rules (16–18). Given the rich structure of social knowledge that humans possess about each other (19,20), it is plausible to assume that this structure is represented and updated during social learning.

No study to date has investigated how social knowledge modulates updating in adolescents with ASD. As such, it remains unclear whether adolescents represent social knowledge when making social inferences and whether this knowledge guides learning. This is a particularly important question given that adolescents with ASD experience more negative social interactions than typically developing (TD) adolescents and an increased risk for comorbidities and emotional maladjustment (21,22).

Computational modeling together with neuroimaging may provide a model for social inferences of adolescents with ASD and important insights into how such inferences are encoded in brain activity. The few studies that have described social

SEE COMMENTARY ON PAGE 760

decisions of adults with high autistic traits or ASD with the use of computational models have found that these individuals represent others' mental states less when making social inferences (23,24). It remains unclear if this difference between typical and autistic individuals is accentuated in adolescence, a unique developmental period characterized by asymmetries between social and general cognition (25,26).

In the present study, we aimed to expand our previous approach of studying social learning in adolescence (9) by investigating how social knowledge structures shape learning. Ideal generalizations for learning should rely on less rigid representations, which afford flexible (re)use. In our previous study, we annotated categories of preferences by hand such that each stimulus was positioned at a specific place of the hierarchy (e.g., a banana is in the subcategory fruit and the category food). We then characterized social learning with computational models that prescribed learning about others' preferences within these predefined categories. More specifically, participants were assumed to learn based on prediction errors (PEs), the difference between initial ratings and subsequent feedback, and use their own preferences to varying degrees. This study extends our previous approach. Here, we constructed additional computational models that rely on peer preferences and scaled PEs based on similarities between preferences. These models assume that individuals represent rich associations between preferences for objects within and between categories. We investigated whether social inferences in TD adolescents and in adolescents with ASD rely on social knowledge representations of varying complexity. To this end, we combined computational modeling with brain imaging to investigate how adolescents make inferences about what individuals from their peer group like and dislike. Participants received trial-by-trial feedback about the peers' actual preference, which they could use to update inferences on similar subsequent items. In addition, to assess whether participants relied on own preferences, we asked participants for their own preferences after scanning. We also collected information about preference structures in a large, independent sample of adolescents. These metrics allowed us to compare how TD adolescents and adolescents with ASD use prior knowledge about themselves and others when making preference inferences.

We hypothesized that individuals with ASD would differ from TD adolescents in the extent to which they rely on prior knowledge about peers, i.e., represent peer preference structures, and how much they update predictions based on social feedback. We expected that differences in representing preference structures and prediction error encoding would scale with differences in neural encoding of social predictions in the MPFC.

METHODS AND MATERIALS

The study consisted of 2 assessments, an online preference survey in a large sample of adolescents (N = 99; 55 female; age, mean ± SD = 15.7 ± 1.4 years; age range = 11–18 years) to establish preferences of the adolescent population, and a functional magnetic resonance imaging (fMRI) experiment in a separate group of TD adolescents (n = 26; 11 female; age, mean ± SD = 13.7 ± 2.5 years; age range = 9–18 years) and adolescents with ASD (n = 20; 12 female; age, mean ± SD = 14.8 ± 2.8 years; age range = 10–20 years). Detailed information on the sample is provided in the Supplement and in Table 1. The experiments have been introduced in our previous report (9), which included 24 of 26 TD adolescents from the current sample. Briefly, the online survey contained pictures of activities, fashion, and food items and a short demographic questionnaire. Survey participants rated how much they liked each item on a 10-point Likert-type scale ranging from 1 (not at all) to 10 (very much). The fMRI experiment was carried out using the same items. Participants were asked to infer the preferences of 3 people from their peer group and were presented with trial-by-trial feedback about the others' actual preference ratings. The rationale behind choosing multiple profiles was to increase the task's ecological validity by increasing the generalizability of observed learning patterns to multiple distinguishable preference profiles. After the fMRI experiment, participants provided their own preferences for the items outside the scanner. They were also asked with open-ended questions to describe the individuals based on what they had learned from the task. Refer to the Supplement for a detailed task description and to Figure 1 for examples. The experimental procedures were conducted in compliance with the standards established by the university's institutional review board and the Declaration of Helsinki.

Behavioral Data Analysis

Our main hypothesis was that learning strategies differ between TD and ASD groups. We hypothesized that TD adolescents would rely on sophisticated knowledge structures (i.e., preference similarities) and PEs, the difference between initial ratings and subsequent feedback, when making preference inferences about peers. We expected these differences in cognitive strategies to be reflected in the amount of PE reduction over the course of the task, whereby TD adolescents would reduce PEs over the course of the task more rapidly than adolescents with ASD. We also expected that learning strategies of TD adolescents would result in more holistic

Table 1. Demographics and Symptom Characteristics

	TD (n = 26)		ASD (n = 20)		
	n	Median	n	Median	p Value
Sex, Female	11		9		.855
Age, Years	26	14.17	20	15.75	.137
DAS—IQ	23	107.5	17	112	.895
SRS II	15	13	18	84	<.001
ADOS		–	20	10	–

The p values were computed based on independent-samples median test for non-normally distributed variables of ordinal and interval scales and using the χ^2 test for variables with nominal scale.

ADOS, Autism Diagnostic Observation Schedule; DAS, Differential Abilities Scale; SRS, Social Responsive Scale.

representations of preference profiles, leading to greater generalizations across items and item categories. These hypotheses were directly tested in model-free and model-based analyses, as detailed below.

To specify individual differences and the developmental trajectory of social learning in adolescents with ASD, we investigated linear and quadratic relationships between demographic variables (e.g., age as decimal values, autistic symptomatology) and PEs and model-based variables (e.g., learning rates from winning model). Given the growing literature on sex differences in ASD, sex was entered as a covariate in these analyses. We assessed whether variables were normally distributed by means of the Kolmogorov-Smirnov test. Nonparametric statistical tests were performed for non-normally distributed variables.

Model-Free Behavioral Analysis. Participants could use the person's feedback on previous items to inform preference inferences for upcoming items, thereby reducing PEs within item categories and beyond. For the model-free analyses, we tested participants' overall accuracy in estimating the other persons by assessing PEs independent of the direction of deviation (positive or negative). We thus defined PEs as the absolute difference between participants' ratings and the feedback they subsequently received. Note that for the model-based analysis described in the next section, PEs are used to adjust upcoming ratings up or down, and are therefore defined as the signed difference between rating and feedback.

To directly test the notion of generalized learning, we also asked participants to provide short descriptions of the persons after the scanner session. We expected TD adolescents to give more detailed descriptions of the persons, mentioning predefined categories and personality traits more frequently in their open-answer descriptions of people after the task (see the Supplement for the rating approach).

Computational Modeling. To test whether learning strategies differ between ASD and TD participants, we devised and formally tested computational models that contain prior knowledge of varying complexity (i.e., own preferences, average peer preferences, preference similarities) and additional reinforcement-learning (RL) components (see the schematic depiction of main models in Figure 1).

Computational Model Space: Main Models

Previously, our main model space consisted of 3 models that assume that participants adjust their estimated ratings of another person (ERs):

1) By performing a simple linear transformation of their own preferences (OP) to predict the preferences of the other persons (model 1: No-Learning).

$$ER = b_0 + b_1 \times OP$$

2) According to a variant of the Rescorla-Wagner RL rule (model 2: RL-Ratings). This model comprises the RL rule, which adjusts expected ratings of the other person's preference, ER_{t+1}, on the basis of the participant's current estimate, ER_t, and the PE. The PE is the difference between this current estimate and the current feedback, F_t, which is the other person's actual preference rating. The prediction error is weighted be the learning rate α, which is a free parameter.

$$ER_{t+1} = ER_t + \alpha PE_t$$

with

$$PE_t = F_t - ER_t$$

3) By using a weighted combination of RL and their own preferences to predict the others' preferences (OP; model 3: Combination [Comb]).

$$ER_{t+1} = \gamma(ER_t + \alpha PE_t) + (1-\gamma)OP_{t+1}$$

Here, we added 4 analogous models described below that additionally include simple and more sophisticated prior knowledge about preferences of peers. They scale PEs by preference similarities and/or substitute participants' own preferences with average population preferences from the adolescent survey.

In model 4 (Simple Prior) participants are assumed to perform a linear transformation from population averages to participants' ratings on the task. Participants base their estimated ratings (ER) of another person on the mean preference rating (MP) for the item in question.

$$ER = b_0 + b_1 \times MP$$

Model 5 (Combination Simple Prior [Comb Simple Prior]) is analogous to the Comb model (model 3). Instead of own preferences this model contains a trade-off between RL and the average population MP.

$$ER_{t+1} = \gamma(ER_t + \alpha PE_t) + (1-\gamma)MP_{t+1}$$

Model 6 (Similarity-RL) represents a more sophisticated version of the RL ratings (model 2). An important difference between this and model 2 is that for this model the integration of PEs is scaled by the preference similarity for the current item i and all subsequent items $i \in I$ that participants are going to see. The preference similarity is conceptualized as the correlation r between preferences of the independent population of 100 adolescents for the current item i and all subsequent items in the task set I.

$$ER_{t+1} = ER_t + \alpha PE_t\, r(i, I)$$

with

$$PE_t = F_t - ER_t$$

Model 7 (Similarity-Combination [Similarity Comb]) represents a more sophisticated version of the Comb model (model 3). Similar to the logic of the Comb model, this model assumes that participants' ERs for the other person rely on a weighted combination of the RL rule with similarity weights (model 5: RL-Ratings) and average population MP to predict the others' preferences.

$$ER_{t+1} = \gamma[ER_t + \alpha PE_t\, r(i, I)] + (1-\gamma)MP_{t+1}$$

Additional models and initialization procedures are included in the Supplement.

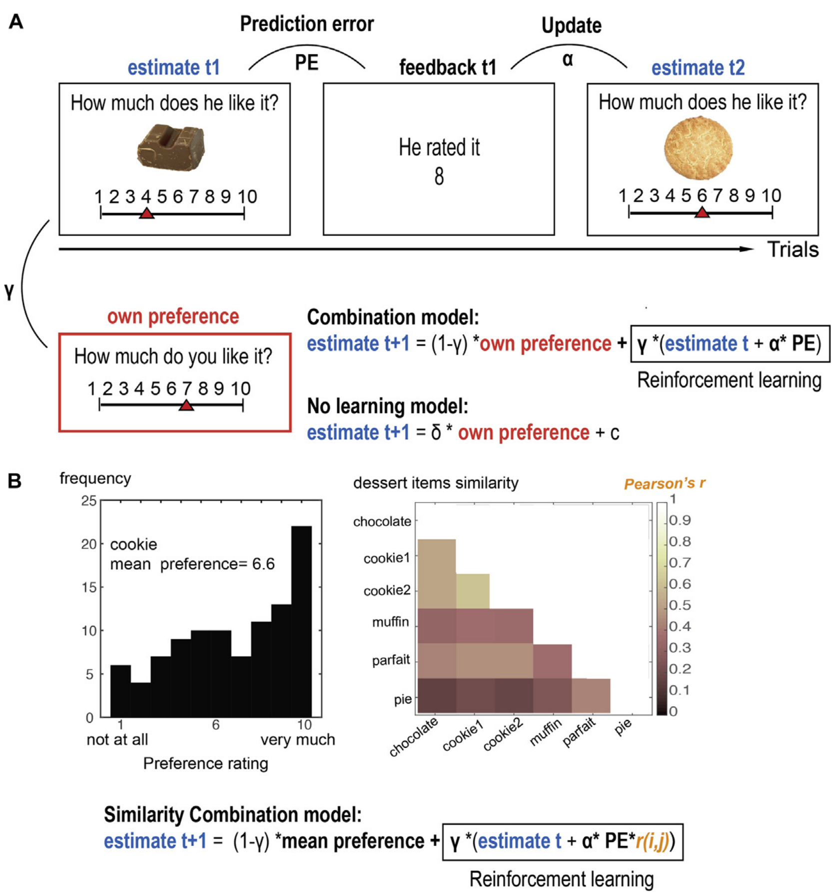

Figure 1. Experimental task and computational models. **(A)** Example item for the preference task and computational models derived from task and own preference information. **(B)** Example distribution and mean rating for one item from the adolescent preference survey with $N = 99$ adolescents for the item cookies. Correlation matrices depict the relationship among ratings of participants' own preferences for different classes of items within an example subcategory (desserts). Mean preferences and preference similarity is used in the Similarity Combination model. PE, prediction error.

Model Estimation and Comparison

We used linear least squares estimation to determine best-fitting model parameters. Optimization used a nonlinear Nelder-Mead simplex search algorithm (implemented in the MATLAB [The MathWorks, Inc., Natick, MA] function *fminsearch*) to minimize the sum of squared errors of prediction over all trials for each participant. We constrained parameter values to sensible a priori defined ranges (i.e., α, γ to a range of 0 to 1). For each model and each participant, we approximated model evidence by calculating the Bayesian information criterion (BIC). Taking a more comprehensive approach to arbitrate between models, we also report the Akaike information criterion (AIC) and AIC weights (27) (Supplement).

Model Validation Procedures

We have implemented a model validation approach in line with recent guidelines for computational modeling (28) (Supplement). In brief, we find that our models are distinguishable from each other and the model parameters recoverable. The predictions of the various models were low to moderately correlated, which ensures that each model captures unique aspects of participants' behavior.

fMRI Data Analysis: Statistical Model

For details on fMRI acquisition and analysis, refer to the Supplement. Two general linear models (GLMs) were set up for each participant and run to investigate how brain activity was modulated by task variables on a trial-by-trial basis (model-free: GLM 1) and by variables derived from the winning behavioral model (model-based: GLM 2). GLM 1 included the following parametric regressors: 1) a regressor for the rating phase, 2) a regressor for the feedback phase, 3) rating numbers to account for variance explained by the presentation of numbers in the rating phases, 4) feedback numbers to control for variance explained by the presentation of numbers in the feedback phase, 5) model-free PEs in feedback phases, and 6) own preferences in rating phases. This regressor tests whether participants encode own preferences based on the No-Learning strategy. GLM 2 included the following parametric regressors: 1) a regressor for the entire rating phase, 2) one regressor for the entire feedback phase, 3) feedback numbers as parametric regressors in feedback phases to account for variance explained by the presentation of numbers only, 4) model-predicted ratings, entered in rating phases, and 5) model-predicted PEs, entered in feedback phases (see Table 2 for correlations between regressors).

RESULTS

Differences in Social Inferences Between Typical and Autistic Adolescents

We tested group differences in task performance with 2 different metrics: unsigned model-free PEs (the numerical differences between estimates and feedback) and open-answer descriptions of preference profiles. Average accuracy, i.e., average absolute model-free Pes, did not differentiate between TD adolescents and adolescents with ASD. Median model-free PEs were on average equally high (median for TD = 2.49; median for ASD = 2.49; H-test on median PEs: $\chi^2_{45} = 0.07$, $p = .790$) and reduced over time to the same degree in both groups ($t_{44} = 0.38$, $p = .705$). Correlations between PEs and demographic variables are reported in the Supplement.

Open-answer descriptions of peers in question revealed higher levels of generalization in the TD group. TD adolescents mentioned predefined categories ($\chi^2_1 = 22.6$, $p < .001$) and personality traits ($\chi^2_1 = 3.52$, $p = .004$) more often in their descriptions than adolescents with ASD. There were no group differences in the average number of words participants wrote in each group ($t_{42} = 0.39$, $p = .696$). Proportions of reported predefined categories and personality inferences for both groups are depicted in Figure 2A.

Cognitive Strategies Underlying Social Inferences in Typical and Autistic Adolescents

In this study, we extended the social inference model introduced previously (9). In the previous study, we found that the Comb model including RL and own preferences (model 3) best described preference inferences of TD subjects. Here, we expanded our previous model space by models that additionally assume knowledge about peer preference structures. By comparing these models with previous models, we found that TD adolescents rely on knowledge about peer preferences. More specifically, preference predictions in the TD group were best captured by the Similarity Comb model. AIC and BIC values were smallest for this model, and the AIC weight showed that this model had the maximal relative likelihood of 1 (Figure 2B). The Similarity Comb model assumes that PEs are scaled by item similarity on a trial-by-trial basis (see Figure 3A for a visualization of PE updating). Note that model fits for each participant in the TD and ASD groups are provided in the supplemental section (Figure S3 and S4). Parameters of the Similarity Comb model were uncorrelated in the TD group, confirming the necessity of both parameters ($r = -.267$, $p = .187$) (Figure 3B).

In contrast, the ASD group was more heterogeneous. According to the BIC, the simple No-Learning model provided the best fit. The model relies solely on the participant's own preferences to predict those of others. According to the AIC and AIC weights, the simple Comb model provided a better fit to the data (Figure 2C). When investigating the distribution of free parameters for the Comb model, however, we found that adolescents with ASD had very low learning rates and relatively high trade-off parameters (γ). The trade-off parameter quantifies how much participants rely on their own preferences versus RL. According to the parameter estimates, ASD participants did not update information based on task feedback (Figure 3C, D). This finding thus corroborated the model comparison using BIC. The intercept and slope parameters for the No-Learning model in the ASD group were highly correlated ($r = -.911$, $p < .001$): the higher the intercept, the lower the slope parameter. The parameters of the Comb model were also negatively correlated ($r = -.462$, $p < .05$), indicating that participants with lower learning rates had higher trade-off parameters (distributions of model parameters are depicted in Figure 3C, D).

To account for the fact that participants may forget feedback about items that were farther away from a current trial, we tested whether decay models, which incorporate a forgetting parameter, explained the behavior better than nondecay models. The decay models performed worse than the main models in the space, indicating that participants did not forget about past feedback (Supplement; Figure S1).

Individual Differences in Model Parameters and Model Evidence. In line with previous reports of ongoing development in learning (9,29), older TD adolescents with greater IQs had higher learning rates, meaning that they adjusted estimates more quickly based on task feedback (Table S3). This relationship between model parameters and age, cognitive ability, and/or symptom severity was not significant in adolescents with ASD using the No-Learning model. Significant group differences were found only in the relationship between cognitive ability and parameter estimates (Fisher's r-to-z = 1.99, $p = .046$).

Given the heterogeneity of our samples, we explored whether model evidence on an individual level related to demographic and symptom characteristics, with particular focus on participants with ASD. These analyses directly related model evidence to demographic and symptom characteristics in an effort to constrain the heterogeneity of ASD with respect to their social skills. We investigated the relationships between Social Responsiveness Scale (30) and IQ scores with BIC values of the winning models—No-Learning and Similarity Comb—in the TD and ASD groups. The Social Responsiveness Scale is a measure of social adjustment with a particular focus on ASD symptomatology. Analyses were corrected for multiple comparisons using false discovery rate (FDR) (31).

The Similarity Comb model evidence was significantly related to social adjustment. We found a significant correlation between Social Responsiveness Scale scores and BIC values for the Similarity Comb model across groups ($\rho = .43$; $p_{\text{FDR corrected}} = .048$) and in ASD adolescents only ($\rho = .62$; $p_{\text{FDR corrected}} = .034$): the lower the model evidence (the larger the BIC values), the more social deficits (Figure S2).

Neural Activity Scaled With Task Variables and Model Predictions

The neuroimaging results corroborated group differences in cognitive strategies. Adolescents with ASD, who relied on their own preferences to predict those of others, exhibited a stronger representation of own preferences in the angular gyrus extending into the precuneus cortex during rating phases, compared with TD adolescents. Similarly, model predictions derived from the winning computational model of TD adolescents were encoded in brain activity during the task.

Table 2. Correlations Between Parametric Regressors in Model-Free and Model-Based Analyses

	Regressors in Model-Free Analysis				Regressors in Model-Based Analysis		
	PE	Feedback	Self Pref		PE	Rating	Feedback
PE	–	−.120[a]	.088	PE	–	.091[a]	.039
Feedback	–.082[b]	–	.191	Rating	.119[b]	–	.355

Feedback, feedback number; PE, prediction error; Rating, model-estimated rating; Self Pref, preference for self.
[a]Adolescents with autism spectrum disorder.
[b]Typically developing adolescents.

Specifically, model-derived PEs from the winning Similarity Comb model were represented in the right caudate and putamen of TD adolescents during feedback phases (Figure 4; Tables S4 and S5).

DISCUSSION

This study investigated how adolescents with and without autism build knowledge representations of their peers. We tested whether adolescents rely on preexisting social knowledge to make preference inferences and how their knowledge is updated during learning about a specific peer. Based on preference profiles from an independent sample of adolescents, we devised computational models that combine knowledge about peers at varying levels of complexity with PE updating. TD adolescents relied on a combination of average population preferences and PE updating scaled by the similarity between current and upcoming items to optimize preference inferences about peers. The ASD group was more heterogeneous. Model comparisons with BIC and AIC metrics indicated different winning models: The No-Learning model won based on BIC, and the Comb model was the winning model based on AIC metrics. Both models, however, rely on participants' own preferences to estimate preferences for another person. While the Comb model assumes that participants rely on RL in addition to their own preferences, the fitted parameter estimates clearly indicated that adolescents with ASD did not update their estimates based on PEs. Predictions of the winning computational models were corroborated by our neuroimaging results. In the TD group, model-derived PEs scaled with brain activity in the putamen. In ASD adolescents, own preference ratings were more closely related to activity in the angular gyrus extending into the precuneus than in TD adolescents.

Social Learning Through Prediction Errors. TD adolescents and adolescents with ASD were asked to infer preferences of peers and could update these predictions based on trial-by-trial feedback. TD adolescents were able to build more generalized representations of the peers in question compared with ASD adolescents. In open-ended questions after the experiment, participants generalized from information about specific items to broader item categories and personality inferences. Our results are in line with a vast body of literature showing that individuals with ASD have core deficits extracting general mental state inferences from social observations (32,33). TD and ASD participants did not differ in the magnitude of PEs or in PE changes over time. The relationship

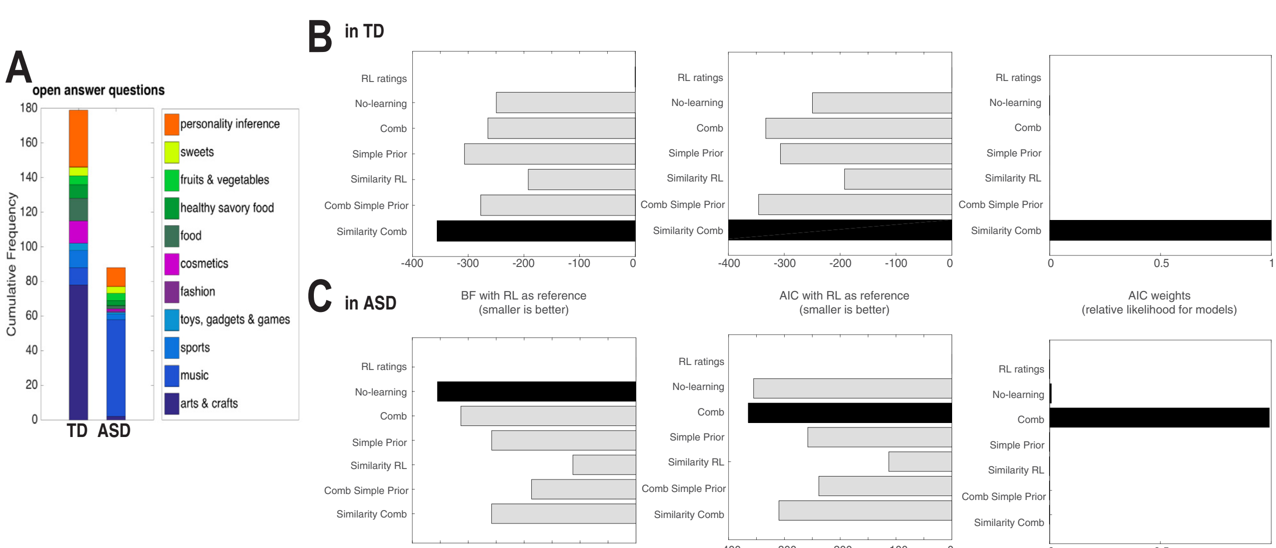

Figure 2. Group differences in social learning. **(A)** Number of predefined categories and personality inferences. Within-group model comparisons using Bayes factor (BF), Akaike information criterion (AIC), and relative likelihoods for model-based Akaike weights in **(B)** typically developing (TD) adolescents and **(C)** adolescents with autism spectrum disorder (ASD). Comb, combination; RL, reinforcement learning.

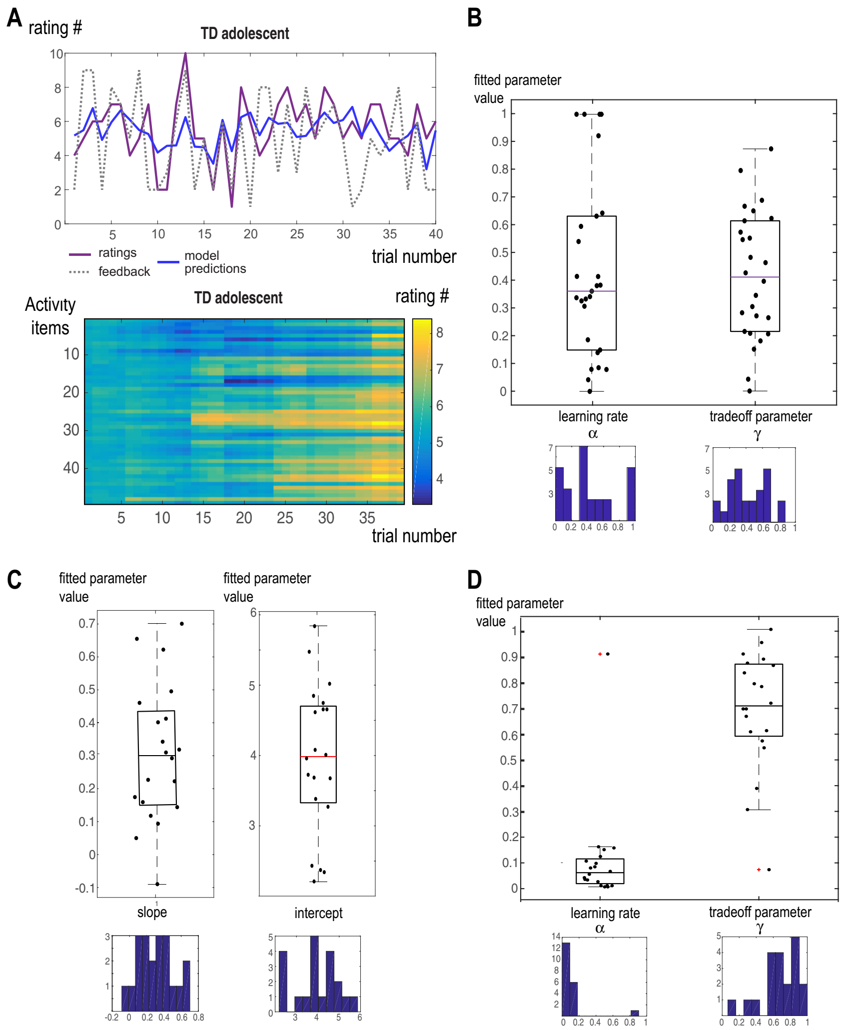

Figure 3. Estimates of winning models in the typically developing (TD) and autism spectrum disorder (ASD) groups. **(A)** Model fit and predictions in one TD adolescent during the task. Lower panel depicts trial-by-trial prediction error updating scaled by item similarity for all activity items simultaneously. At the beginning of the task, estimates rely on item similarities only. Over the course of the task, participants use prediction errors to update all items scaled by their similarity respectively. **(B)** Parameter estimates from the Similarity Combination model in TD adolescents. **(C)** Parameter estimated from the No-Learning model in the ASD group. **(D)** Parameter estimates of the Combination model in the ASD group.

between PEs and demographic variables, however, differed between groups. PEs were significantly related to age in the TD group—with mid-adolescents having larger PEs than early and late adolescents. This nonlinear trajectory of social development is a well-replicated finding. Adolescence is a unique period for social development; mid-adolescents have higher prediction errors (34) and lower learning rates than younger and older peers, which means that they learn more slowly from social feedback (9,29). We did not find such ongoing social development in adolescents with ASD. Instead, adolescents with ASD with higher IQs had overall lower PEs on the social inference task. Previous studies also found that language and academic skills improve social adjustment of adolescents with autism (35,36), while for TD adolescents, social and cognitive skills are relatively independent from them (37,38).

Social Knowledge Representation and Updating. The main aim of this study was to expand our previous computational modeling approach by investigating how social knowledge structures shape learning. Previous studies have shown that knowledge representations are activated during learning (39,40). Our study makes an important contribution to this literature by specifying how social knowledge structures are represented and updated during learning through task-based feedback. The computational modeling approach revealed that choices of TD adolescents rely on average peer preferences and PEs scaled by preference similarities. If, for instance, a participant previously learned about their peer's preference for bananas and now has to judge how much the peer likes apples, the participant would rely on the average population preference for apples and on how much the peer liked bananas before, given the degree to which preferences for bananas and apples are related.

In contrast, the models that best described inferences of adolescents with ASD were simpler. They assumed that participants rely on their own preferences for the item at hand (an inference about how much the peer likes apples relies on how much the participant likes apples). ASD participants seemed to differ from TD adolescents in 2 crucial ways: First, they did not rely on preexisting knowledge about population preferences,

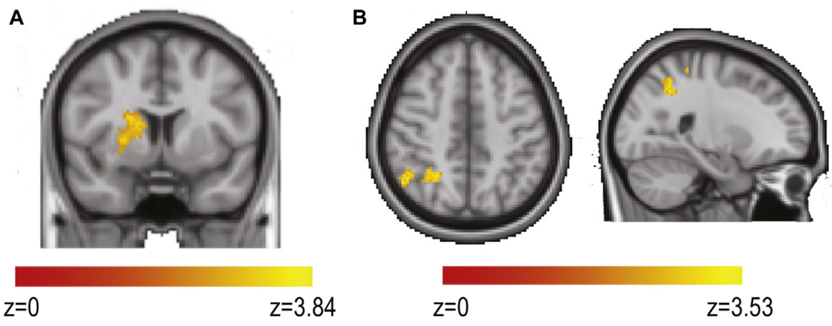

Figure 4. Parametric modulation of brain activity by computational model variables during the social inference task. **(A)** Brain activity scaled with prediction errors predicted by the winning model in typically developing adolescents (the Similarity Combination model). **(B)** Brain activity correlated more strongly with own preferences in the autism spectrum disorder vs. typically developing group. This supports the behavioral findings that adolescents with autism spectrum disorder rely more on their own preferences.

and second, learning rates of the Comb model indicated minimal PE updating from feedback about the person's actual preferences. Failure to encode social PEs and to adjust ratings, respectively, may account for the observed social rigidity of individuals with ASD (41) and explain their lack of preexisting social knowledge structures (42). Our findings are in line with 2 recent accounts of ASD, which posit that individuals with ASD experience deficits in calculating decision variables and/or representing the statistical structure of the environment (24,41–44). When making social inferences, adolescents with ASD used less informative knowledge (i.e., themselves vs. population preferences) and updated information less based on feedback.

Both groups, but particularly the ASD group, showed a large degree of variability in model evidence values. To explore whether this variability could be clinically meaningful, we investigated the relationships of BIC scores for the winning models—Similarity Comb and No-Learning—with demographic variables across groups and for ASD participants only. We found that participants who were better fitted by the more sophisticated Similarity Comb model had higher social skills. This significant relationship across groups and in the ASD group only strengthens the link between social knowledge representation and social adjustment.

Neural Mechanisms Underlying Social Inferences. The computational models that best described participants' behavior were validated by our neuroimaging results. In TD adolescents, PEs estimated by the winning model were encoded in the putamen and caudate, parts of the dorsal striatum. Adolescents show increased activity of these regions when computing decision values, and this has been typically linked to heightened dopaminergic PE responsivity and to increased reward-seeking tendencies of adolescents (26,45). Unlike the ventral striatum, which has been consistently implicated in the computations of model-free prediction errors (46,47), the dorsal striatum, particularly the putamen, has been more closely related to model-based learning, for instance, to encoding model-based prediction errors and stable environmental structure that aids learning (48,49). Our results are in line with these aforementioned studies. The model-based PE encoded in the putamen differs from a simple reward-PE in that it is continuously scaled by preference similarity information.

Model-free PEs scaled with activity of the MPFC, a region that has been extensively implicated in encoding and updating others' mental states (50–52). The discrepancy between regions encoding model-based and model-free PEs highlights the difference between these decision variables: Model-free errors are conceptualized as the discrepancy or absolute differences between participants' ratings and the subsequent feedback. Model-based PEs are signed error values that adjust model estimates up and down. Model estimates are not the actual rating of the participant but reflect population means and the similarity structure used by the model. Dopaminergic neurons in the dorsal striatum may be sensitive to this signed error term, which contains the valence of the preference information (peer likes something more or less than predicted) and similarity structure that can inform learning.

Unlike TD adolescents who encoded both model-based and model-free PEs in neural activity, we did not find neural encoding of PEs in adolescents with ASD. ASD participants encoded their own preferences more strongly in the angular gyrus extending into the precuneus cortex compared with TD adolescents, corroborating their reliance on their own preferences when making inferences about peers during the task. The angular gyrus and precuneus have been identified as integrative regions that store and represents conceptual knowledge and support attention to relevant information (53,54). The precuneus, in particular, has been repeatedly linked to social deficits of individuals with ASD (55,56) and changes in precuneus activity scale with improvements in social cognition (57). To firmly establish the roles of these regions in encoding decision variables during social learning of individuals with ASD, this finding has to be replicated in future studies with larger sample sizes and more restricted age ranges. The large age range in our sample may have precluded us from detecting systematic group differences in brain regions that undergo significant development during adolescence, such as the prefrontal cortex (58).

In a similar vein, the present study lacks a representative sample of ASD preference profiles to rule out the idea that adolescents with ASD rely on representations about peers with ASD. However, we did not actually tell any of the participants whether the preferences came from adolescents with or without autism. Furthermore, given that ASD is a rather rare neurodevelopmental disorder, it is questionable whether adolescents with ASD have the opportunity to build elaborate social knowledge structures about peers with ASD. Future studies should assess whether ASD preference profiles differ from those of TD individuals and, if so, should probe whether individuals with ASD use knowledge representations of peers with ASD during social learning.

Furthermore, in this study we could not investigate whether participants represent preference similarities with a multivariate approach, such as representational similarity analysis (59), because participants received feedback during the task. Feedback changes the representation of items and their similarities. To test the neural representation of the preference similarity matrix, future studies should add pure estimation trials to the task, in which participants cannot learn because they do not receive feedback.

Finally, in accordance with current recommendations to take a second-person approach to social neuroscience (60), future studies should explore whether the learning strategies identified here generalize to learning during on-line social interactions. This generalization would more readily allow translation of the insights gained here into clinical and educational practice.

In summary, this study sheds light on adolescent social development in several important ways. First, it extends our previous account of social learning in adolescence by showing that typical adolescents represent sophisticated social knowledge during learning that is continuously updated through input from the environment. Second, to our knowledge, this is the first study that specifies the cognitive strategies underlying social learning about others' preferences in autism. Finally, our results provide evidence for the usefulness of a neurocomputational approach in describing social development and social deficits of individuals with neuropsychiatric disorders.

ACKNOWLEDGMENTS AND DISCLOSURES

This research was supported by the Hilibrand Fellowship for Autism Research at the Yale Child Study Center.

We thank Koen Frolichs for assisting us with data analysis.

This article was published as a preprint on bioRxiv: https://doi.org/10.1101/850552.

DL is a cofounder of Neurogazer. The other authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Center for Translational Developmental Neuroscience (GR, AD, KAP), Yale Child Study Center, Yale University, New Haven, Connecticut; Institute for Systems Neuroscience (CWK), University Medical Center Hamburg-Eppendorf, Hamburg, Germany; Autism and Neurodevelopmental Disorders Institute (GR) and Department of Psychological and Brain Sciences (GR), George Washington University, Washington, DC; Zanvyl Krieger Mind/Brain Institute (DL), Department of Neuroscience, Department of Psychological and Brain Sciences, Johns Hopkins University, Maryland; and Department of Neurology (KAP), University of Virginia, Charlottesville, Virginia.

Address correspondence to Gabriela Rosenblau, Ph.D., at grosenblau@gwu.edu.

Received Feb 14, 2020; revised Jun 1, 2020; accepted Jun 6, 2020.

Supplementary material cited in this article is available online at https://doi.org/10.1016/j.bpsc.2020.07.002.

REFERENCES

1. Blakemore SJ, Mills KL (2014): Is adolescence a sensitive period for sociocultural processing? Annu Rev Psychol 65:187–207.
2. Steinberg L (2008): A social neuroscience perspective on adolescent risk-taking. Dev Rev 28:78–106.
3. Burnett S, Sebastian C, Cohen Kadosh K, Blakemore SJ (2011): The social brain in adolescence: Evidence from functional magnetic resonance imaging and behavioural studies. Neurosci Biobehav Rev 35:1654–1664.
4. Casey BJ, Jones RM, Hare TA (2008): The adolescent brain. Ann N Y Acad Sci 1124:111–126.
5. Dumontheil I, Burgess PW, Blakemore SJ (2008): Development of rostral prefrontal cortex and cognitive and behavioural disorders. Dev Med Child Neurol 50:168–181.
6. Luciana M (2013): Adolescent brain development in normality and psychopathology. Dev Psychopathol 25:1325–1345.
7. Giedd JN, Blumenthal J, Jeffries NO, Castellanos FX, Liu H, Zijdenbos A, *et al.* (1999): Brain development during childhood and adolescence: A longitudinal MRI study. Nat Neurosci 2:861–863.
8. Somerville LH (2013): Special issue on the teenage brain: Sensitivity to social evaluation. Curr Dir Psychol Sci 22:121–127.
9. Rosenblau G, Korn CW, Pelphrey KA (2018): A computational account of optimizing social predictions reveals that adolescents are conservative learners in social contexts. J Neurosci 38:974–988.
10. Somerville LH, Jones RM, Ruberry EJ, Dyke JP, Glover G, Casey BJ (2013): The medial prefrontal cortex and the emergence of self-conscious emotion in adolescence. Psychol Sci 24:1554–1562.
11. Pelphrey KA, Carter EJ (2008): Charting the typical and atypical development of the social brain. Dev Psychopathol 20:1081–1102.
12. Jones CRG, Simonoff E, Baird G, Pickles A, Marsden AJS, Tregay J, *et al.* (2018): The association between theory of mind, executive function, and the symptoms of autism spectrum disorder. Autism Res 11:95–109.
13. Elsabbagh M, Johnson MH (2010): Getting answers from babies about autism. Trends Cogn Sci 14:81–87.
14. Dawson G, Meltzoff AN, Osterling J, Rinaldi J, Brown E (1998): Children with autism fail to orient to naturally occurring social stimuli. J Autism Dev Disord 28:479–485.
15. Klin A, Jones W, Schultz R, Volkmar F (2003): The enactive mind, or from actions to cognition: Lessons from autism. Philos Trans R Soc B Biol Sci 358:345–360.
16. Reber AS (1989): More thoughts on the unconscious: Reply to Brody and to Lewicki and Hill. J Exp Psychol Gen 118:242–244.
17. Seger CA (1995): Efficiency and conceptual fluency as independent mechanisms in implicit learning. Diss Abstr Int Sect B Sci Eng 55:5107.
18. Berry DC, Broadbent DE (1984): On the relationship between task performance and associated verbalizable knowledge. Q J Exp Psychol Sect A 36:209–231.
19. Fiske ST, Taylor SE (2016): Social Cognition: From Brains to Culture, 3rd ed. Sage Publications.
20. Leary MR (2007): Motivational and emotional aspects of the self. Annu Rev Psychol 58:317–344.
21. Corbett BA, Simon D (2013): Adolescence, stress and cortisol in autism spectrum disorders. OA Autism 1:2.
22. Seltzer MM, Krauss MW, Shattuck PT, Orsmond G, Swe A, Lord C (2003): The symptoms of autism spectrum disorders in adolescence and adulthood. J Autism Dev Disord 33:565–581.
23. Yoshida W, Dziobek I, Kliemann D, Heekeren HR, Friston KJ, Dolan RJ (2010): Cooperation and heterogeneity of the autistic mind. J Neurosci 30:8815–8818.
24. Sevgi M, Diaconescu AO, Tittgemeyer M, Schilbach L (2016): Social Bayes: Using Bayesian modeling to study autistic trait–related differences in social cognition. Biol Psychiatry 80:112–119.
25. Blakemore SJ, Choudhury S (2006): Development of the adolescent brain: Implications for executive function and social cognition. J Child Psychol Psychiatry 47:296–312.
26. Blakemore SJ, Robbins TW (2012): Decision-making in the adolescent brain. Nat Neurosci 15:1184–1191.
27. Wagenmakers EJ, Farrell S (2004): AIC model selection using Akaike weights. Psychon Bull Rev 11:192–196.
28. Wilson RC, Collins AG (2019): Ten simple rules for the computational modeling of behavioral data. Elife 8:e49547.

29. Jones RM, Somerville LH, Li J, Ruberry EJ, Powers A, Mehta N, *et al.* (2014): Adolescent-specific patterns of behavior and neural activity during social reinforcement learning. Cogn Affect Behav Neurosci 14:683–697.
30. Constantino JN, Gruber CP (2005): Social Responsiveness Scale (SRS). Western Psychological Services.
31. Storey JD (2002): A direct approach to false discovery rates. J R Stat Soc Ser B Stat Methodol 64:479–498.
32. Yang DY-JY-JJ, Rosenblau G, Keifer C, Pelphrey K a (2015): An integrative neural model of social perception, action observation, and theory of mind. Neurosci Biobehav Rev 51:263–275.
33. Rosenblau G, Kliemann D, Heekeren HR, Dziobek I (2015): Approximating implicit and explicit mentalizing with two naturalistic video-based tasks in typical development and autism spectrum disorder. J Autism Dev Disord 45:953–965.
34. Dumontheil I, Apperly IA, Blakemore SJ (2010): Online usage of theory of mind continues to develop in late adolescence. Dev Sci 13:331–338.
35. Magiati I, Tay XW, Howlin P (2014): Cognitive, language, social and behavioural outcomes in adults with autism spectrum disorders: A systematic review of longitudinal follow-up studies in adulthood. Clin Psychol Rev 34:73–86.
36. Howlin P, Goode S, Hutton J, Rutter M (2004): Adult outcome for children with autism. J Child Psychol Psychiatry Allied Discip 45:212–229.
37. Hu Z, Chan RCK, McAlonan GM (2010): Maturation of social attribution skills in typically developing children: An investigation using the social attribution task. Behav Brain Funct 6:10.
38. Baron-Cohen S, Leslie AM, Frith U (1985): Does the autistic child have a “theory of mind”? Cognition 21:37–46.
39. Garvert MM, Dolan RJ, Behrens TEJ (2017): A map of abstract relational knowledge in the human hippocampal-entorhinal cortex. Elife 6: e17086.
40. Constantinescu AO, O’Reilly JX, Behrens TEJ (2016): Organizing conceptual knowledge in humans with a gridlike code. Science 352(6292):1464–1468.
41. Sinha P, Kjelgaard MM, Gandhi TK, Tsourides K, Cardinaux AL, Pantazis D, *et al.* (2014): Autism as a disorder of prediction. Proc Natl Acad Sci U S A 111:15220–15225.
42. Pellicano E, Burr D (2012): When the world becomes “too real”: A Bayesian explanation of autistic perception. Trends Cogn Sci 16:504–510.
43. Friston KJ, Lawson R, Frith CD (2013): On hyperpriors and hypopriors: Comment on Pellicano and Burr. Trends Cogn Sci 17:1.
44. van de Cruys S, Evers K, van der Hallen R, van Eylen L, Boets B, de-Wit L, Wagemans J (2014): Precise minds in uncertain worlds: Predictive coding in autism. Psychol Rev 121:649–675.
45. Cohen JR, Asarnow RF, Sabb FW, Bilder RM, Bookheimer SY, Knowlton BJ, Poldrack RA (2010): A unique adolescent response to reward prediction errors. Nat Neurosci 13:669–671.
46. Garrison J, Erdeniz B, Done J (2013): Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. Neurosci Biobehav Rev 37:1297–1310.
47. Gläscher J, Daw N, Dayan P, O’Doherty JP (2010): States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66:585–595.
48. Haruno M, Kawato M (2006): Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. J Neurophysiol 95:948–959.
49. Kunimatsu J, Maeda K, Hikosaka O (2019): The caudal part of putamen represents the historical object value information. J Neurosci 39:1709–1719.
50. Hare TA, Camerer CF, Knoepfle DT, O’Doherty JP, Rangel A (2010): Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. J Neurosci 30:583–590.
51. Garvert MM, Moutoussis M, Kurth-Nelson Z, Behrens TE, Dolan RJ (2015): Learning-induced plasticity in medial prefrontal cortex predicts preference malleability. Neuron 85:418–428.
52. Behrens TE, Hunt LT, Woolrich MW, Rushworth MF (2008): Associative learning of social value. Nature 456:245–249.
53. Seghier ML (2013): The angular gyrus: Multiple functions and multiple subdivisions. Neuroscientist 19:43–61.
54. Joiner J, Piva M, Turrin C, Chang SWC (2017): Social learning through prediction error in the brain. NPJ Sci Learn 2:8.
55. Liu J, Yao L, Zhang W, Xiao Y, Liu L, Gao X, *et al.* (2017): Gray matter abnormalities in pediatric autism spectrum disorder: A meta-analysis with signed differential mapping. Eur Child Adolesc Psychiatry 26:933–945.
56. Jack A, Morris JP (2014): Neocerebellar contributions to social perception in adolescents with autism spectrum disorder. Dev Cogn Neurosci 10:77–92.
57. Rosenblau G, O’Connell G, Heekeren HR, Dziobek I (2019): Neurobiological mechanisms of social cognition treatment in high-functioning adults with autism spectrum disorder. Psychol Med 1–11.
58. Blakemore SJ (2008): The social brain in adolescence. Nat Rev Neurosci 9:267–277.
59. Kriegeskorte N, Kievit RA (2013): Representational geometry: Integrating cognition, computation, and the brain. Trends Cogn Sci 17:401–412.
60. Redcay E, Schilbach L (2019): Using second-person neuroscience to elucidate the mechanisms of social interaction. Nat Rev Neurosci 20:495–505.