

RESEARCH ARTICLE

Revealing dichotomous prior biases in social anxiety through a social prism model

Yuxi Wang^{1,2}, Qianqian Ju^{3*}, Renhe Jia^{1,4}, Minghao Yuan^{1,4}, Yujia Peng^{1,2,5*}

1 School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Brain-Computer Interface and Mental Health Modulation, Peking University, Beijing, P.R.China, **2** State Key Laboratory of General Artificial Intelligence, Beijing Institute for General Artificial Intelligence, Beijing, P.R.China, **3** School of Psychology, Shanghai University of Sport, Shanghai, P.R.China, **4** Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing, P.R.China, **5** Institute for Artificial Intelligence, Peking University, Beijing, P.R.China

* juqianqian@sus.edu.cn (QJ), yujia_peng@pku.edu.cn (YP)



OPEN ACCESS

Citation: Wang Y, Ju Q, Jia R, Yuan M, Peng Y (2026) Revealing dichotomous prior biases in social anxiety through a social prism model. PLoS Comput Biol 22(7): e1014509. <https://doi.org/10.1371/journal.pcbi.1014509>

Editor: Haojiang Ying, Soochow University, CHINA

Received: February 28, 2026

Accepted: June 29, 2026

Published: July 27, 2026

Copyright: © 2026 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: Data availability Source data for Figs. 2–5, as well as for Supplementary Figures and Table, are available at <https://github.com/yuxiwang670/dichotomous-prior-biases-in-SA> and https://osf.io/txjef/overview?view_only=e5fdac197ab74d8eac95a417276688b8.

Abstract

Social situations can be overwhelming for some people, triggering avoidance and social anxiety (SA). However, it remains unknown why identical situations lead to different interpretations. Here, we developed a Social Prism Model to systematically address mechanisms underlying atypical social cognition in SA within a Bayesian cognitive framework. Through eight social valence judgment experiments (N = 541), we demonstrated that social anxiety may be shaped not by how sensory evidence is processed, but by robust dichotomous prior biases depending on social cues. The dichotomous prior biases indicated two parallel mental shortcuts that predispose individuals toward social fear, where negative prior biases were associated with an intolerance to uncertainty and a fear of social evaluations, and over-positive prior biases were associated with negative social learning. Our Bayesian simulations further demonstrated how variations in prior expectations parameters can give rise to biased social judgments, providing mechanistic support for the proposed framework. Collectively, these findings showed that the negative interpretation of the social situations can be understood within the Bayesian framework, highlighting a key role of prior expectation in shaping human social cognition.

Author summary

Social situations can be overwhelming for some individuals, yet it remains unclear why the same scene triggers very different responses across people. Through a series of empirical human studies, we systematically decomposed social scenes into social cues of faces, bodies, animacy, emotion, interaction, and ambiguity. We showed that biased social cognition in social anxiety may arise from conflicting expectations about the social world, rather than from difficulty processing social cues. People with high social anxiety traits hold negative prior

Funding: This study was partly supported by the National Science and Technology Major Project SQ2022AAA010294 (to Y.P.), the National Natural Science Foundation of China (32471151, 32200854 to Y.P.), the Young Elite Scientists Sponsorship Program by China Association for Science and Technology 2021QNRC00 (to Y.P.), the Postdoctoral Fellowship Program of CPSF GZC20230105 (to Q.J.), the research project of Shanghai University of Sport 2026STD002 (to Q.J.), and the New Faculty Research Startup Funding of Shanghai University of Sports (to Q.J.). Q.J. received salary support from the Postdoctoral Fellowship Program of CPSF. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

biases associated with intolerance to uncertainty and a fear of social evaluations, and at the same time, overly positive prior biases that reinforce mal-adaptive social learning. Together, our work offers a theory-driven framework for understanding heterogeneity in biased social cognition among social anxiety individuals, with implications for more targeted mental health interventions.

Introduction

Social situations can be overwhelming for some people, triggering avoidance and social anxiety (SA). However, it remains unknown why identical situations lead to different interpretations. Atypical social cognition in SA often manifests as misinterpreting social cues [1,2]. For instance, when giving a presentation, a casual frown from the audience can be perceived by some individuals as dissatisfaction with a failed presentation. In contrast, others may consider the signal random and remain unaffected. Despite the existence of numerous studies on atypical social cognition in SA, the findings have been inconsistent. Some studies have reported negative interpretation biases in SA [3,4], while others found no such bias [5] or even a positive bias [6]. These inconsistencies point to the need for a unified computational framework to systematically characterize the cognitive mechanisms underlying atypical social cognition in SA. To address this issue, social cognition can be conceptualized as involving reciprocal interactions between prior expectations and signal accumulation of social cues, which together shape our social reactions and interpretations [7,8] (Fig 1A). Prior expectations guide the interpretation of information by integrating past experiences with current cues, while the signal accumulation process incorporates key social information from actual scenarios [8]. Therefore, biases in social cognition may stem from distortions in either or both prior expectations and signal accumulation. For example, a negatively biased prior expectation can skew social interpretation [9], and over-accumulation of negative signals can similarly lead to negative interpretations [10]. Here, we propose the importance of disentangling biased prior expectations from mal-adaptive signal accumulation, as well as the specific contributions of social cues. Clarifying these mechanisms is critical for advancing both theoretical models and clinical interventions.

Computational modeling provides a principled solution to disentangle prior expectation from signal accumulation and to quantify their contributions. In the current study, we adopted the Hierarchical Drift-Diffusion Model (HDDM, Fig 2A) [11,12] to model social cognition with four key model parameters: the initial bias (z), the drift rate (v), the threshold boundary (a), and the non-decision time (t_0) [13]. The initial bias reflects a starting tendency in information processing, with a higher z indicating a predisposition to the upper-bound choice [14]. The drift rate reflects the efficiency of signal accumulation, with a greater v indicating faster signal accumulation and higher sensitivity [15]. Here, prior expectations and signal accumulation were mapped onto the initial bias and drift rate, respectively. The remaining two HDDM parameters, a and t_0 , indicate the style of decision-making (impulsive or deliberate) and the time that was irrelevant to the decision-making process, respectively [16].

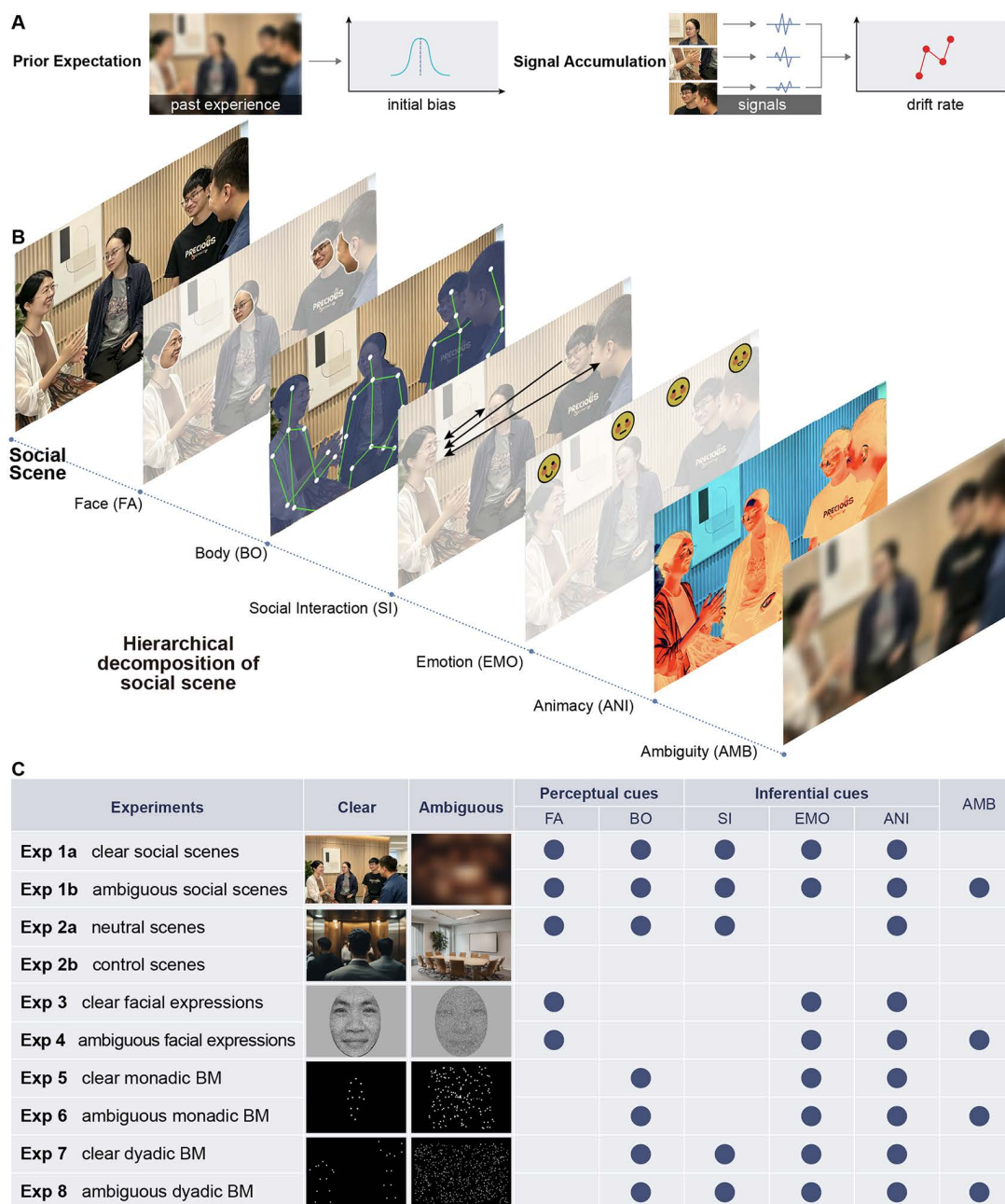


Fig 1. The decomposition of social scenes via a Social Prism framework. (A) An illustration of processes of prior expectation and signal accumulation underlying social cognition of social scenarios. (B) The Social Prism systematically deconstructs the complex social scenarios into hierarchical components of social cues and ambiguity. Social cues were further decomposed into perceptual cues (face, body) and inferential cues (social interaction, emotion, animacy). (C) Stimuli used in all the experiments and the mapping of experimental stimuli onto social cues decomposed by the Social Prism framework, with circles indicating the presence of specific social cues in each experiment. Facial and bodily cues were defined by the presence of a face or a body, respectively. Animacy was indicated by the presence of something being a living thing; social interaction by scenes involving engagement between two or more agents; and emotion by the presence of affective valence. Ambiguity was introduced through controlled visual noise or image blurring. BM=biological motion, Exp=Experiment.

<https://doi.org/10.1371/journal.pcbi.1014509.g001>

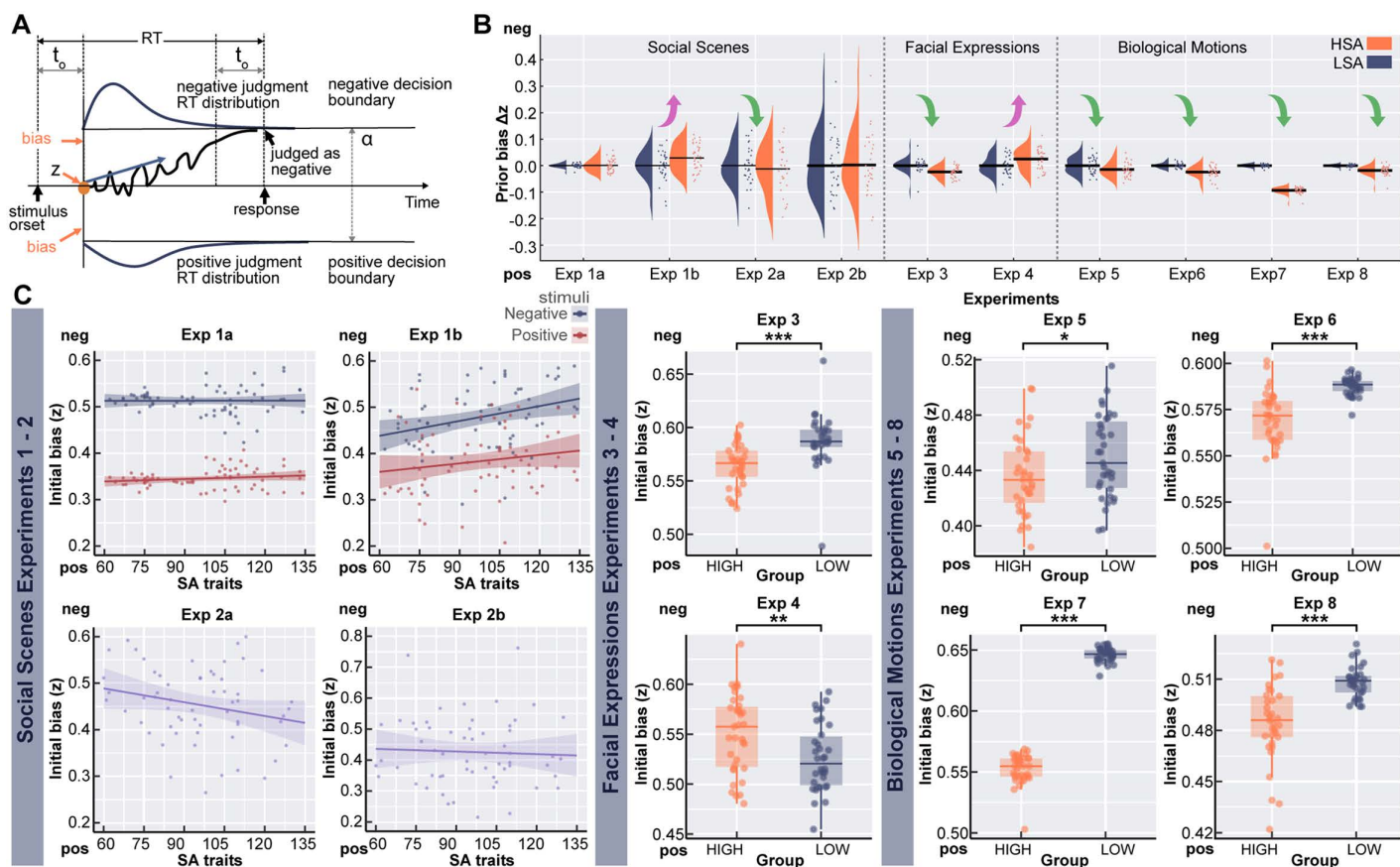


Fig 2. HDDM model parameters and associations with SA traits. (A) An illustration of the HDDM model and parameters. HDDM model depicts human decision-making as an evidence accumulation process that eventually reaches one of two boundaries, yielding four key parameters of initial bias (z), drift rate (v), threshold boundary (a), and non-decision time (t_0). (B) Prior biases (Δz) across eight experiments, reflecting individual deviations in prior expectations relative to the reference ($\Delta z = 0$, dashed line), the mean z -values in the LSA groups). Dots represent individual participants. Colored violin plots display the distribution of prior biases for HSA (orange) and LSA (blue) groups. Solid horizontal black lines indicate the group means of prior bias for HSA and LSA, respectively. The LSA groups were consistently centered on the no-bias reference ($y = 0$), while HSA groups demonstrated prior biases in several conditions. Pink arrows indicate experiments where HSA participants showed negative prior biases, whereas green arrows denote experiments where HSA participants exhibited positive prior biases. (C) Associations between SA traits and initial biases and drift rates across eight experiments. Solid lines indicate best-fitting linear regressions; shaded areas represent 95% confidence intervals. Box-plots display the median (central line) and interquartile range (box). Scatters denote individual participant data points. Statistical significance: $*p < .05$, $**p < .01$, $***p < .001$ (Bonferroni-corrected). RT=Reaction time; neg=negative, pos=positive; BM=biological motion; HSA=high social anxiety, LSA=low social anxiety.

<https://doi.org/10.1371/journal.pcbi.1014509.g002>

Furthermore, daily social interactions involve complex and heterogeneous social cues, such as multi-agent interactions, body postures, faces, and emotional signals, that may influence the social cognition of SA in different ways [17]. However, previous studies have primarily focused on isolated social cues such as facial expressions [5,18], which may oversimplify the representation of social information, especially when SA was found to associate with an avoidance of direct gaze at faces [19]. Thus, to understand atypical social cognition and the origin of social fear, it is necessary to decompose the contributions of different social cues. To systematically decompose complex social scenes into social cues while preserving the overall composite of social context, we propose a Social Prism framework in analogy to how a physical prism disperses white light into its component wavelengths (Fig 1B). The framework decomposes social scenes into perceptual cues and inferential cues, where perceptual cues consist of face (FA) and body (BO) to be directly observed, and

inferential cues consist of animacy (ANI) [20], social interaction (SI), and emotion (EMO) to be inferred from perceptual cues. Furthermore, ambiguity (AMB) is incorporated into the framework to capture situations where social information is uncertain or unclear. Under clear conditions, both prior expectations and sensory accumulation contribute to social cognition. Under ambiguous conditions, however, visual signals become less reliable, and social judgements depend more heavily on prior expectations, amplifying the influence of existing biases in prior expectations on social cognition.

The current study investigates whether biased social cognition originates in distorted prior expectations or altered accumulation of social evidence. We hypothesize that biased priors are a core computational mechanism underlying social-cognitive biases, and that distinct social cues within the Social Prism framework are linked to separable prior representations, each contributing uniquely to social cognition while being integrated within a unified computational framework.

Results

Eight behavioral experiments of social valence judgements were conducted to examine the social cognition in participants with high and low SA traits. For experimental stimuli, as shown in Fig 1C, we adopted social scenes (Experiments 1–2), isolated facial expressions (Experiments 3–4), and biological motions (Experiments 5–8). In all the experiments, participants were asked to judge the valence of the presented stimuli, for a social scene as “positive” or “negative”, or for facial expressions and biological motions as “happy” or “angry”. Participants were asked to respond as quickly and accurately as possible. Responses and reaction times (RTs) were recorded for analysis. Participants were enrolled based on the Chinese Social Anxiety Questionnaire for Adults (CSAQ-A) [21,22], which measures the SA traits of individuals. For detailed participant information, see Table 1.

Individual differences in behavioral judgements of social valence

On the behavioral level (i.e., responses and RTs; S2 Fig), results did not show a consistent pattern of group differences across tasks (S2 Fig). For responses, we found a significant positive association between SA traits and the ratio of negative choices only in the negative ambiguous social scenes (Exp. 1b, $\beta = 0.238$, $R^2 = 0.057$, $p = .044$), but not for any other experiments ($ps > 0.05$). For RTs, in the clear monadic biological motion experiment (Exp. 5), a repeated-measures analysis of variance (ANOVA) revealed a significant interaction between group and stimulus valence, $F(1,68) = 4.651$, $p = .035$, $\eta_p^2 = 0.064$. Post hoc comparisons showed that both the high social anxiety (HSA) group and the low social anxiety (LSA) group recognized happy emotions more slowly than angry emotions, with this effect being more pronounced in the HSA group. No other significant group $\times$ stimulus valence interactions were observed in the remaining experiments.

HDDM parameters reveal initial biases in SA

Among the HDDM parameters, results showed robust associations between SA traits and initial biases (z) across experiments (Fig 2C) except for clear social scenes (Exp. 1a) and control scenes (Exp. 2b). Specifically, SA traits were positively correlated with z in negative ($\beta = 0.233$, $p = .005$) and positive ambiguous social scenes (Exp. 1b, $\beta = 0.205$, $p = .013$), and negatively correlated with z in neutral scenes (Exp. 2a, $\beta = -0.258$, $p = .043$). For facial expressions, two-sample t -tests revealed that HSA exhibited a significantly more positive z in clear facial expressions than LSA (Exp. 3, $t(61) = 3.916$, $p < .001$, $d = 1.012$), but significantly more negative z in ambiguous facial expressions (Exp. 4, $t(67) = 2.813$, $p = .006$, $d = 0.676$). In biological motion tasks, HSA consistently demonstrated more positive z than LSA (Exp. 5, $t(68) = 2.600$, $p = .040$, $d = 0.500$; Exp. 6, $t(62) = -5.610$, $p < .001$, $d = -1.370$; Exp. 7, $t(68) = 40.826$, $p < .001$, $d = 9.760$; Exp. 8, $t(68) = -5.900$, $p < .001$, $d = -1.410$). These results suggest that initial biases may be key contributors to atypical social cognition in individuals with high SA traits. However, instead of homogeneous contributions, initial biases contributed differently to the biased social cognition depending on the combination of social cues. Besides z , we did not find robust group differences for other HDDM parameters (v , a , t_0) across experiments (S5 Material in S1 Text, S3, S4 Figs).

Table 1. Demographic characteristics of participants included in each experiment.

Experiments	Group	Gender		Age		Social Anxiety Trait	
		Female	Male	<i>M</i> (<i>SD</i>)	Min, Max	<i>M</i> (<i>SD</i>)	Min, Max
Exp. 1(<i>N</i> =73)	HSA (<i>N</i> =45)	28 (62.22%)	17 (37.78%)	21.40 (2.24)	18,27	114.00(12.00)	95,139
	LSA (<i>N</i> =28)	19 (67.86%)	9 (32.14%)	21.10 (2.27)	18,26	76.70(7.64)	61,92
Exp. 2(<i>N</i> =62)	HSA (<i>N</i> =32)	20 (62.50%)	12 (37.50%)	22.40 (3.34)	18,30	110.00 (10.70)	95,130
	LSA (<i>N</i> =30)	20 (66.67%)	10 (33.33%)	23.10 (3.60)	18,34	76.60 (12.30)	52,94
Exp. 3(<i>N</i> =63)	HSA (<i>N</i> =34)	22 (64.70%)	12 (35.30%)	21.85 (2.86)	18,28	122.62 (6.20)	115,139
	LSA (<i>N</i> =29)	20 (68.96%)	9 (31.04%)	21.55 (3.25)	18,32	62.96 (9.07)	41,73
Exp. 4(<i>N</i> =69)	HSA (<i>N</i> =35)	27 (77.14%)	8 (22.86%)	21.20 (2.52)	18,29	123.20 (5.46)	110,137
	LSA (<i>N</i> =34)	23 (67.65%)	11 (32.35%)	22.60 (3.70)	18,34	67.94 (8.50)	47,80
Exp. 5(<i>N</i> =70)	HSA (<i>N</i> =35)	23 (65.71%)	12 (34.29%)	21.91 (2.84)	18,28	122.48 (6.18)	115,139
	LSA (<i>N</i> =35)	23 (65.71%)	12 (34.29%)	21.65 (3.12)	18,32	65.60 (10.13)	41,80
Exp. 6(<i>N</i> =64)	HSA (<i>N</i> =33)	19 (57.58%)	14 (42.42%)	21.91 (2.92)	18,30	122.54 (7.34)	110,144
	LSA (<i>N</i> =31)	17 (54.84%)	14 (45.16%)	21.51 (2.33)	18,26	71.00 (8.37)	51,81
Exp. 7(<i>N</i> =70)	HSA (<i>N</i> =35)	27 (77.14%)	8 (22.86%)	21.20 (2.52)	18,29	123.20 (5.46)	110,137
	LSA (<i>N</i> =35)	24 (68.57%)	11 (31.43%)	22.60 (3.70)	18,34	67.94 (8.50)	47,80
Exp. 8(<i>N</i> =70)	HSA (<i>N</i> =35)	20 (57.14%)	15 (42.86%)	21.91 (2.92)	18,30	122.54 (7.34)	110,144
	LSA (<i>N</i> =35)	20 (57.14%)	15 (42.86%)	21.51 (2.33)	18,26	71.00 (8.37)	51,81

Note. While most experiments recruited independent samples, some participants contributed to more than one experiment conducted within the same session (Exp. 3 & 5, Exp. 4 & 7, Exp. 6 & 8). Participant counts reflect final samples after exclusion. Across all experiments, trials with log-transformed response times exceeding ± 3 SD from the participant mean were removed. In Experiments 5–8, participants with overall accuracy <80% were excluded, and trials with response times <300 ms or exceeding stimulus duration were removed. HSA=high social anxiety; LSA=low social anxiety; *M*=Mean; *SD*=Standard Deviation.

<https://doi.org/10.1371/journal.pcbi.1014509.t001>

The Social Prism Model revealed dichotomous prior biases in SA

We quantified prior bias (Δz) as each participant's deviation in z from the mean of the LSA group within the same experiment (Fig 2B; Eq. 2). We then applied a linear regression approach, incrementally adding social cue predictors to assess whether dimensions of the Social Prism explained individual differences in prior bias. The full model, which incorporated all six social cues and is hereafter referred to as the Social Prism Model (Fig 3C, Eq. 3), provided the best fit to the data (highest $-\Delta AIC$; $R^2=0.139$; Fig 3A).

Regression and simple slope analyses revealed that all social cues of the Social Prism were significantly associated with prior biases in SA (Fig 3B). However, the directions of associations were not homogeneous (Fig 3D). Specifically, high SA trait was associated with **over-negative** prior biases in response to FA ($\beta=1.27 \times 10^{-3}$, $p<.001$) and AMB ($\beta=7.08 \times 10^{-3}$, $p<.001$), but **over-positive** prior biases in response to ANI ($\beta=-1.04 \times 10^{-3}$, $p<.001$), SI ($\beta=-1.25 \times 10^{-3}$, $p<.001$), EMO ($\beta=-1.04 \times 10^{-3}$, $p<.001$), and BO ($\beta=2.02 \times 10^{-3}$, $p<.001$). Together, these findings demonstrate dichotomous prior biases that depend on social cues among individuals with high SA traits, highlighting the cue-specific nature of prior biases in SA.

We further evaluated the performance of predicting SA traits through machine learning (ML, Fig 4). The ML model took three feature sets: behavioral features (e.g., RTs), latent features derived from HDDM parameters (e.g., z and v), and their combined features, as training inputs to predict SA traits. Results showed that latent features consistently outperformed behavioral features in predicting SA traits across experiments, except Experiments 2, 4, and 5. For instance, in Exp. 1b, latent features achieved a $M_{EVS} = 0.627$ ($SD_{EVS} = 0.199$) compared to $M_{EVS} = 0.316$ ($SD_{EVS} = 0.288$) for behavioral features ($t(38) = 3.967$, $p<.001$). Moreover, across all experiments except Exp. 6, models using latent features

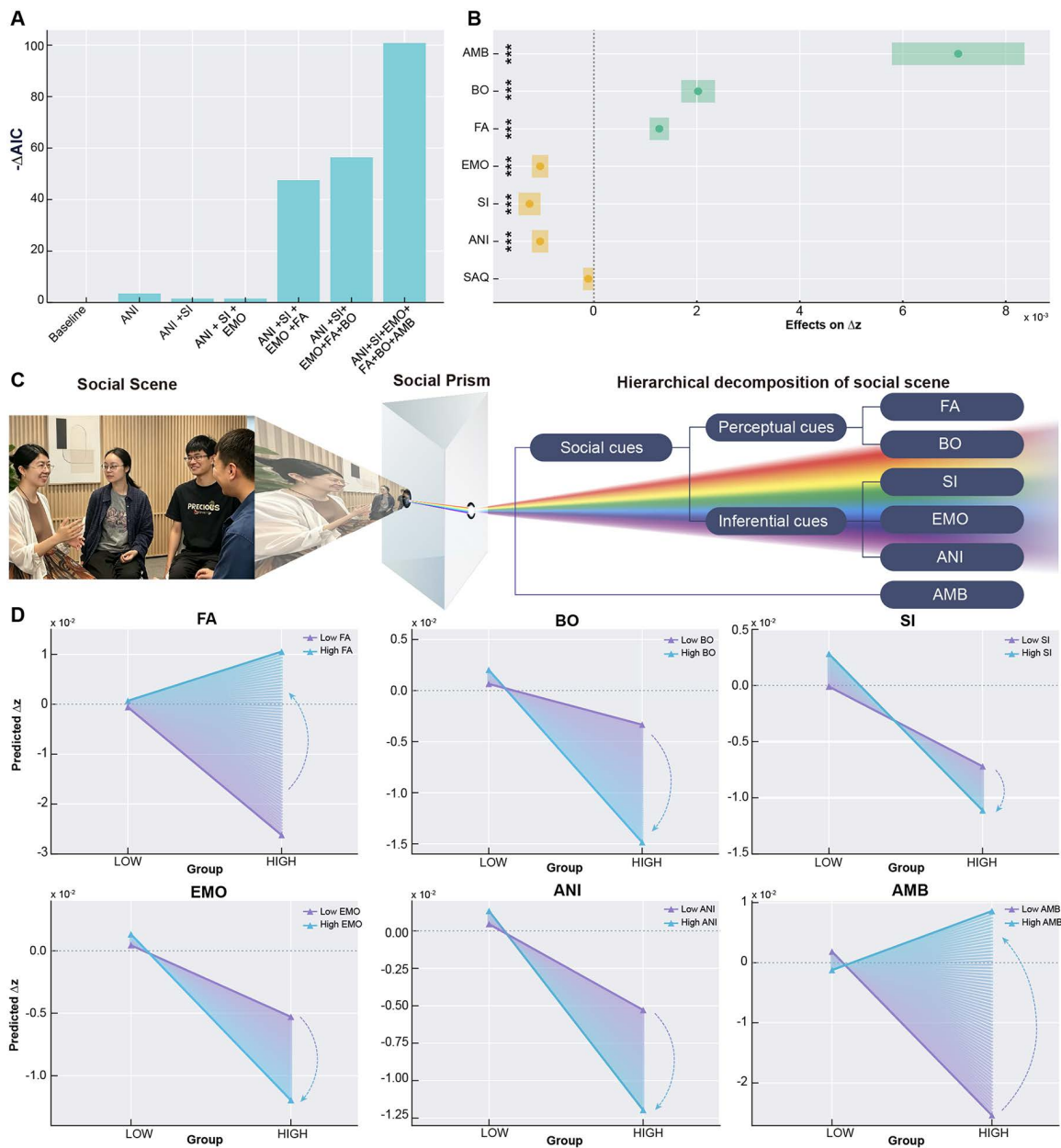


Fig 3. The Social Prism Model and predictive evaluation of social cognition in SA. (A) Linear regression model comparison. $-\Delta AIC$ for incremental addition of social cues in the Social Prism as predictors, including ANI, SI, EMO, FA, BO, and AMB. The full model, referred to as the Social Prism Model, yielded the largest $-\Delta AIC$, indicating the best overall model fit. (B) Estimated effects of each social cue on Δz in the Social Prism Model. Dots indicate mean effect estimates; error bars represent 95% confidence intervals; $***p < .001$. (C) The Social Prism systematically deconstructs the complex social scenarios into hierarchical components of social cues and ambiguity. Social cues were further decomposed into perceptual cues (face, body) and inferential cues (social interaction, emotion, animacy). (D) Interaction plots illustrating the estimated Δz values across levels of SA (LSA vs. HSA) and each social cue. FA=face, AMB=ambiguity, ANI=animacy, SI=social interaction, EMO=emotion, BO=body.

<https://doi.org/10.1371/journal.pcbi.1014509.g003>

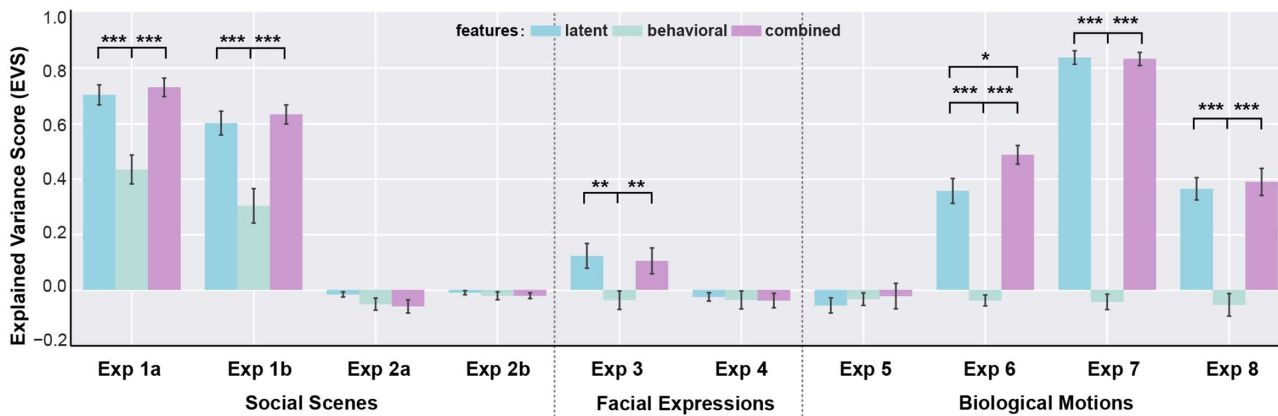


Fig 4. Machine learning results. Model performance comparison across feature sets using explained variance scores (EVS). Models were trained using behavioral features (green), latent features from HDDM (blue), and combined features (purple). The error bar indicates the standard error of EVS. Statistical significance: * $p < .05$, ** $p < .01$, *** $p < .001$.

<https://doi.org/10.1371/journal.pcbi.1014509.g004>

performed comparably to the combined feature models (Exp. 6, $t(38) = 2.327$, $p = .025$). For full results, see S7 Material in [S1 Text](#) and [S1 Table](#).

Bayesian simulation captures biased social cognition

To simulate biased social cognition arising from biased prior expectations, we developed a Bayesian generative model to simulate and reproduce valence judgements in ambiguous social scenes (Exp. 1b). This approach aimed to replicate the behavioral findings that HSA individuals were associated with increased ratios of negative judgements. Specifically, we estimated likelihood parameters by projecting CNN-extracted valence features into a low-dimensional space, where a K-nearest neighbors model identified cluster centroids and computed parameters based on distances to ambiguous images ([Fig 5A](#)). These likelihoods, together with prior expectations, were incorporated into a one-step Bayesian inference model to simulate participants' binary decisions ([Fig 5B](#)).

The Bayesian simulations successfully reproduced key behavioral patterns observed in Experiment 1b. At the individual level, the simulated trial-wise responses demonstrated high accuracy when comparing model-generated responses against empirical responses as a reference ([Fig 5C](#), $M \pm SD = 0.674 \pm 0.067$), significantly above chance level ($t(72) = 22.092$, $p < .001$). Simulated and actual negative response rates were highly correlated across participants ([Fig 5D](#), $r = 0.840$, $p < .001$), demonstrating the model's ability to capture participants' overall response pattern. SA traits were significantly associated with simulated negative response rates for negative stimuli ($\beta = 0.250$, $p = .033$), but not for positive stimuli ($\beta = 0.174$, $p = .140$) ([Fig 5E](#)). Together, the results provide converging evidence that biased social cognition in SA can be accounted for by a Bayesian inference framework.

Discussion

In the present study, the Social Prism Model revealed dichotomous prior biases that may drive biased social cognition in the context of SA. SA individuals showed over-negative prior biases toward facial cues and ambiguity, but over-positive prior biases in response to animacy, emotion, social interaction, and bodily cues. This dichotomous pattern may suggest a cue-specific rather than generalized fear of social scenarios underlying atypical social cognition. In particular, the negative biases associated with facial expressions may be linked to a fear of social evaluation [2], which may predispose SA individuals to automatically expect negative outcomes in response to facial cues, even in the absence of overt threat.

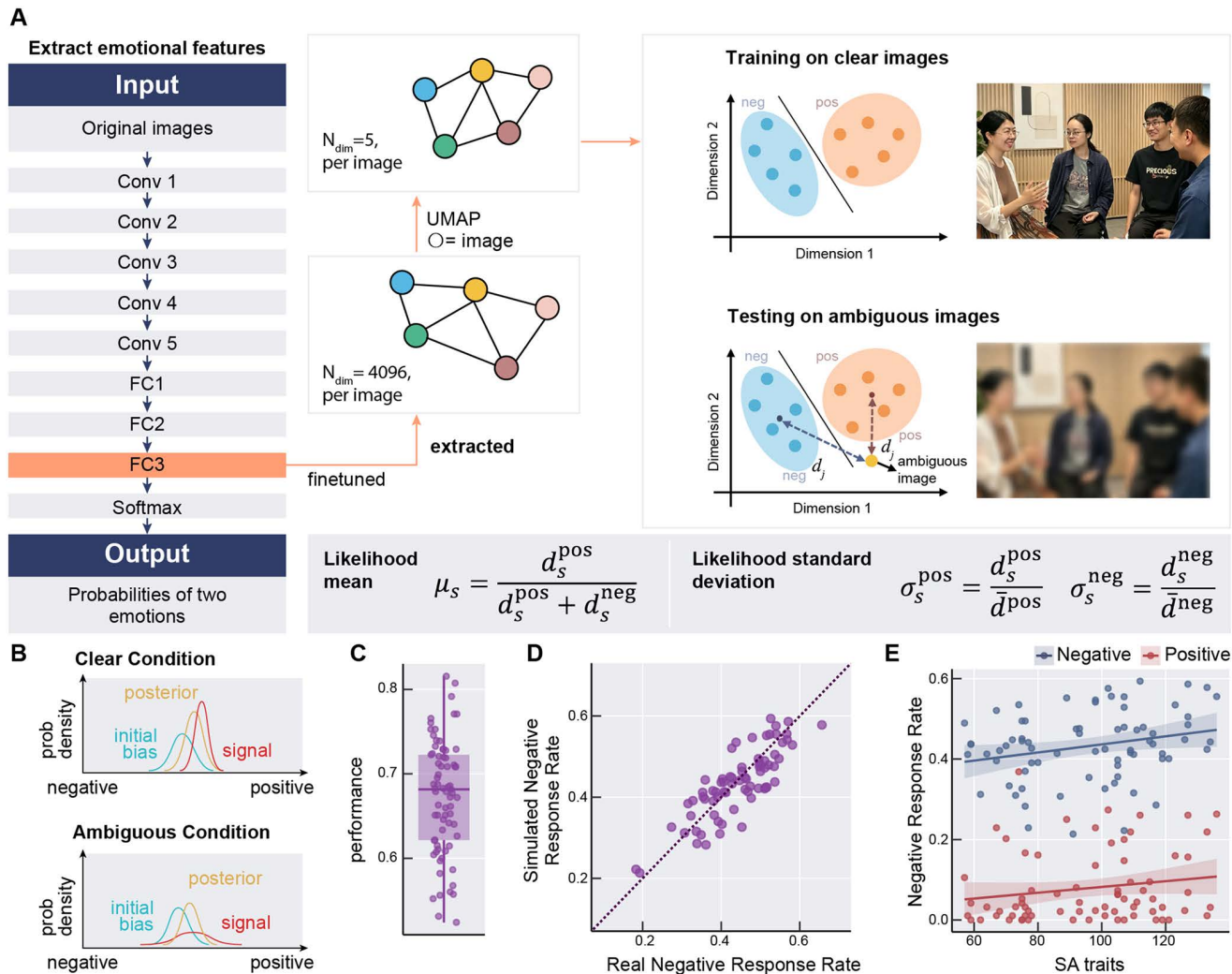


Fig 5. Bayesian simulation for atypical social cognition in SA under ambiguous social scene conditions. (A) Likelihood parameters estimation. Valence features were extracted from social scenes using a fine-tuned convolutional neural network, and projected into a low-dimensional valence space via Uniform Manifold Approximation and Projection (UMAP). A K-nearest neighbors model (KNN) trained on clear images identified the centroids of the positive and negative valence clusters, and the mean and standard deviation of the likelihood were computed based on Minkowski distances from each ambiguous image to the class centroids. (B) A one-step Bayesian inference framework was used to simulate participants' binary decisions based on prior expectations and stimulus-driven likelihoods. For clear condition, prior expectation and signal accumulation both contributed to the social cognition. Under ambiguous condition, the visual signal is no longer reliable, and it heightens reliance on prior expectations during social cognition. For participant i and stimulus s , the prior was derived from the individual-level HDDM-estimated initial bias (z), while the likelihood was determined by the valence features of the stimulus. Simulated decisions were sampled from the resulting posterior distribution, and binarized based on a threshold (see Methods), where 0 indicates a positive judgement and 1 indicates a negative judgement. (C) Accuracy of simulated valence judgements compared to empirical valence judgements across participants. Each dot represents one individual. (D) Simulated negative response rates closely matched actual negative response rates across participants, indicating high generative fidelity of the model. (E) Simulated behavior showed that SA traits were associated with increased negative response rates, particularly to negative ambiguous stimuli, which replicated patterns observed in actual behavioral data.

<https://doi.org/10.1371/journal.pcbi.1014509.g005>

Similarly, the negative biases associated with ambiguity may be associated with intolerance of uncertainty [23–25]. When social scenarios are ambiguous, intolerance of uncertainty in SA may trigger negative automatic thoughts such as expecting social rejection or negative judgements [26–28], thereby leading to an over-negative prior bias.

In contrast, the presence of over-positive prior biases in response to animacy, emotion, social interaction, and bodily cues may reflect atypical social learning that could reinforce strong social fear and avoidance in SA [29,30]. When facing neutral social cues (e.g., a non-smiling face), which may be more common than positive cues in daily life, an over-positive prior may lead to a negative discrepancy between the expectation and reality, contributing to disappointment or perceived rejection. Thus, counter-intuitively, the over-positive prior biases could potentially contribute to negative reinforcement learning in social scenarios for SA individuals. However, future studies are needed to empirically examine the association between prior biases and social learning.

Together, the dichotomous prior biases may help clarify the persistence of atypical social cognition in SA under repeated social exposure [26–28]. Importantly, behavioral responses emerged from the integration of prior expectations and incoming social evidence [31,32]; however, differences across social cue conditions were explained more strongly by variations in prior expectations than by differences in evidence accumulation. At a cognitive level, this finding suggests that cue-specific expectations derived from past social experiences may exert a stronger influence on social interpretations, while still operating in concert with evidence processing during social inference. For potential clinical practice, current evidence may suggest a shift of attention from social skill training to the recalibration of cue-specific prior biases. Specifically, interventions targeting over-negative priors toward facial and ambiguous cues may help reduce mal-adaptive expectations of social evaluation and rejection [33], whereas interventions addressing over-positive priors toward animacy, emotion, social interaction, and bodily cues may help recalibrate unrealistic expectations about social outcomes [34]. Such cue-specific modifications could potentially be integrated into cognitive-behavioral therapy techniques for SA [35–38] such as cognitive restructuring and related techniques [39]. As a complement, third-wave therapies, including mindfulness-based approaches [37,38,40], may align with the proposed framework by reducing the dominance of mal-adaptive priors and increasing tolerance of uncertainty through acceptance, decentering, and cognitive defusion. However, because these approaches primarily target individuals' relationship with their thoughts and experiences, they may not directly modify the content of cue-specific priors and may therefore be insufficient to fully recalibrate prior biases when implemented alone. Future studies are needed to determine whether and how these interventions differentially modify prior expectations, evidence processing, and their integration during social inference.

Importantly, the dichotomous prior biases identified in the current study may help reconcile inconsistent behavioral findings in previous literature on atypical social cognition in the context of SA. For instance, the null effect reported by Jusyte and Schönenberg [5] may reflect the cancellation of opposing effects driven by different social cues: an over-positive prior bias triggered by emotional information and a negative prior bias associated with faces. In contrast, the negative interpretation bias observed by Heuer et al. [3] likely emerged because the stimuli initially presented neutral facial expressions that gradually morphed into emotional ones. In the absence of strong early emotional cues to elicit positive priors, the default negative bias toward faces may dominate and shape the negative interpretation. These examples highlight the importance of disentangling the specific social cues involved in shaping prior expectations. Separately, the absence of signal accumulation effects in our study may be attributable to differences in stimulus modality, as our use of rapid visual social cues compared to the verbal stimuli used in previous work [10].

In addition, Bayesian simulations in the current study accurately reproduced participants' response patterns, supporting a cue-structured account of predictive inference, in which distinct social inputs, such as faces, bodies, or emotional information, may be associated with separable generative models. Within the Bayesian framework, these predictive processes are driven by prior expectations, which guide the interpretation of incoming social information [41–44]. Because the present simulation was implemented only in ambiguous social scenes, which incorporated all cue dimensions represented in the Social Prism framework, it should be regarded as an initial proof-of-concept demonstration. Future studies extending similar computational modeling approaches across additional social cue domains will be important for evaluating the generalizability of this account and providing broader validation of the framework.

At a neural level, the Social Prism Model may be supported by partially distinct neural pathways involving top-down regulation of different categories of social information [45–47]. Specifically, over-negative priors toward facial and ambiguous cues may be linked to threat-related networks involving the amygdala, insula, anterior cingulate cortex, and medial prefrontal cortex [48,49]. In contrast, over-positive priors toward animacy, emotion, social interaction, and bodily cues may rely more heavily on social perception and mentalizing networks, including the superior temporal sulcus, temporo-parietal junction, and medial prefrontal cortex [50,51]. Future studies combining computational modeling with neuroimaging approaches may help determine how cue-specific prior biases are represented across neural systems during social inference. Furthermore, electroencephalography (EEG) and event-related potentials (ERPs) may help characterize the temporal dynamics of social inference and identify when cue-specific prior biases emerge during information processing. Early ERP components such as P1 and N170 may be particularly relevant to the perceptual processing of social cues, whereas later components such as P3 and the late positive potential (LPP) may reflect evaluative processing and belief updating [52–55]. Identifying these temporal dynamics may help clarify how prior biases unfold across processing stages and provide additional constraints for computational models of atypical social cognition.

A few limitations remain for the current study. First, the Social Prism Model is primarily limited to static social images or decontextualized dynamic biological motion stimuli. Incorporating dynamic social scenarios with greater complexity would improve its ecological validity and better capture real-world social inference. Moreover, although the Social Prism Model explained variability in cue-specific prior biases, the origins of these biases remain unclear. Factors operating at broader developmental and contextual levels, such as developmental history [56] and fluctuations in state anxiety [57], may contribute to the formation or modulation of cue-specific priors. Future work should integrate these influences to extend the model. Second, the relatively large effects observed in the biological motion experiments should be interpreted with caution. One possible explanation is that biological motion provides dynamic social information that unfolds over time, potentially increasing the influence of cue-specific prior biases on behavioral responses. Third, the current findings are based on trait-level SA in the general population, rather than clinically diagnosed cases. Due to strong social avoidance and high comorbidity of SA with depression and other anxiety disorders [2], recruiting “pure” SA patients in clinical settings remains challenging. Future studies that validate the model in clinical samples merit attention. Moreover, the identified prior biases were inferred from group-based contrasts defined by established subclinical social anxiety thresholds. Future research should examine whether these effects generalize along continuous dimensions of individual differences. Finally, the current study does not establish the causal role of biased prior expectations in atypical social cognition. Future research could address this limitation by experimentally manipulating prior probabilities [14,58].

In summary, the Social Prism Model provided a unified computational account of social anxiety. By framing biased social processing as a trade-off between cue-specific over-negative and over-positive prior expectations, the model indicated two parallel pathways contributing to the atypical social cognition in the context of SA: fear of negative evaluation resulting from over-negative expectations, and negative social learning resulting from violations of over-positive expectations. Bayesian simulations further demonstrated how individual differences in prior bias parameters shape social cognitive behaviors, providing computational support for the proposed mechanisms. This framework provided a formal foundation for translating core components of cognitive-behavioral models [26,27,59,60] into a predictive processing account of social cognition. More broadly, this work highlighted the promise of computational psychiatry that goes beyond symptomatology [61,62] and uncovered latent cognitive mechanisms, thereby motivating theory-driven interventions.

Methods

Ethics statement

All experiments have been approved by the Ethics Committee of Peking University (#2022-02-10). All participants provided written informed consent before participating in the experiments. Experiments were preregistered on the Open

Science Framework (OSF) before data collection. The full pre-registration is available at https://osf.io/nypaz/overview?view_only=9ce4a968b6f145b5be0b7ca10f5e010f (Experiments 1–2) and https://osf.io/kyw7a/overview?view_only=aa6ea58c555d40f2884eea012ed28646 (Experiments 3–8).

Social anxiety questionnaire

The Chinese Social Anxiety Questionnaire for Adults (CSAQ-A) [21,22] was used to measure the individual levels of SA trait. Total scores range from 30 to 150, with higher scores indicating greater SA trait levels. The normative mean for Chinese young adults was 94.93, and the cut-off scores for the high and low SA groups were 110 and 81, respectively [22]. Group definitions are provided in S1 Material in [S1 Text](#).

Sample size determination

Sample size determinations were based on a combination of a priori power analyses and post hoc sensitivity analyses using G*Power (Version 3.1) [63]. For Experiments 1–2, no prospective power analysis was conducted. Instead, a post hoc sensitivity analysis was performed to estimate the minimum detectable effect size for the linear regression models. To ensure a conservative estimate, we based the analysis on Experiment 2, which had the smallest sample size ($N=62$; Exp. 1: $N=73$). Assuming two predictors (i.e., SA trait scores and stimulus valence), $\alpha=.05$ and 80% power, the analysis indicated sensitivity to effects of Cohen's $f^2 \geq 0.164$, corresponding to a moderate effect size. This suggests that the study had sufficient power to detect moderate effects. For Experiments 3–8, sample size planning was informed by a priori power analysis for a 2 (group) $\times$ 2 (stimulus valence) repeated-measures ANOVA. Assuming an interaction effect of $f=0.19$, $\alpha=.05$, 80% power, a correlation among repeated measures of 0.5, and nonsphericity correction $\epsilon=1$, the analysis indicated that a minimum sample of $N=58$ would be sufficient.

Valence judgement tasks and experiment design

Experiment 1. Clear and ambiguous social scenes. Participants. In Experiment 1, we enrolled 45 HSA (28 females; age 21.40 ± 2.24 years; CSAQ-A 114.00 ± 12.00) and 28 LSA subjects (19 females; age 21.10 ± 2.27 years; CSAQ-A 76.70 ± 7.64). The two groups were matched in age and gender ([Table 1](#)). All participants had normal or corrected-to-normal vision, without any visual impairments. All participants had no current or historical psychiatric or neurological illness. Participants were compensated with ¥40–50.

Stimuli. Social scene images were adopted from the Social Artificial Intelligence Picture System (SAIPS) [64], with half demonstrating positive valence and half negative valence. The valence score of the images ranged from 1 (highly negative) to 9 (highly positive). The formal experiment utilized 48 positive images ($M=7.42$) and 48 negative images ($M=3.85$), as shown in [Fig 1C](#). Ambiguous images were created by applying a 2D Gaussian smoothing kernel ($\sigma=55$) to the clear images using MATLAB's "imgaussfit".

Procedure. Social scenes were presented with a visual angle of 10° on a screen with a resolution of 1920×1080 pixels and a refresh rate of 60 Hz, with a viewing distance of 70 cm. Participants first underwent a practice phase consisting of 4 trials. Each trial began with a fixation cross displayed for 800–1200 ms, followed by the presentation of a social scene. Participants were instructed to quickly and accurately judge whether the perceived valence conveyed by the image was positive or negative by pressing the left (key "F") or right (key "J") button.

The task consisted of 4 blocks, each containing 48 trials (24 positive stimuli and 24 negative stimuli), for a total of 192 trials. No feedback was provided during the task. The first two blocks presented clear images, while the last two involved ambiguous social scenes. Response key assignments were counterbalanced across both blocks and participants. The order of stimulus presentation within each block was randomized.

Experiments 2a and 2b. Neutral scenes and control scenes. Participants. In Experiment 2, we enrolled 32 HSA (20 females; age 22.40 ± 3.34 years; CSAQ-A 110.00 ± 10.70) and 30 LSA subjects (20 females; age 23.10 ± 3.60 years; CSAQ-A 76.60 ± 12.30). The two groups were matched in age and gender (Table 1).

Stimuli and Procedure. A total of 48 neutral social scenes and 48 control scene images were adopted from SAIPS (Fig 1C) [64] with valence scores ranging from 4 to 6 (neutral: $M=4.67$; control: $M=4.91$). The experimental procedure was identical to that of Experiment 1. A total of 96 trials were presented in random order in two blocks, with 48 trials in each block.

Experiments 3–4. Clear and ambiguous facial expressions. Participants. Experiment 3 followed the recruitment process and criteria as in Experiment 1. A total of 34 HSA (22 females; age 21.85 ± 2.86 years; CSAQ-A 122.62 ± 6.20) and 29 LSA participants (20 females; age 21.55 ± 3.25 years; CSAQ-A 62.96 ± 9.07) were enrolled. For Experiment 4, ambiguous face expressions, 35 HSA (27 females; age 21.20 ± 2.52 years; CSAQ-A 123.20 ± 5.46) and 34 LSA participants (23 females; age 22.60 ± 3.70 years; CSAQ-A 67.94 ± 8.50) were enrolled. Both groups were matched on age and gender across two experiments (Table 1). Participants received an average compensation of ¥35 per experiment.

Stimuli. A total of 45 happy and 45 angry facial expressions with balanced gender were adopted from the NimStim Set of Facial Expressions (NimStim) [65] and the Yonsei Face Database (YFace DB) [66]. All images were standardized for luminance, color contrast, and eye alignment, then converted to grayscale. For clear faces in Experiment 3 (Fig 1C), mild Gaussian noise ($M=0$, $SD=0.1$) was added using MATLAB's "imnoise" function to mitigate the ceiling effect. More intense noise ($M=0$, $SD=2$) was applied to create ambiguous faces for Experiment 4 (Fig 1C).

Procedures. Facial expressions were presented with a visual angle of 6.4° [67]. The facial expressions were presented against a background color calculated from the average of the facial image pixels. Participants first went through a practice phase with 10 trials. Each trial started with a 0.5-second fixation cross, followed by the display of facial stimuli. Participants were asked to judge whether the perceived emotion conveyed by the facial expression was happy or angry as quickly and accurately as possible, by pressing left or right buttons. Participants were required to reach an accuracy of 0.9 in practice to proceed to the test. Otherwise, the practice phase was repeated until the criterion was met. The testing phase contained 6 blocks, each with 30 trials, yielding a total of 180 trials, with half angry and half happy stimuli. No feedback was provided during the test. On each trial, if no response was received within the 3-second response window, the trial would be repeated at the end of the test. The sequence of stimulus presentation was randomized for each participant, and button pressing was balanced across participants.

Experiments 5–6. Clear and ambiguous monadic biological motions. Participants. 35 HSA participants (23 females; age 21.91 ± 2.84 years; CSAQ-A 122.48 ± 6.18) and 35 LSA participants (23 females; age 21.65 ± 3.12 years; CSAQ-A 65.60 ± 10.13) were recruited for the clear monadic biological motion experiment (Exp. 5). Ambiguous monadic biological motion experiment (Exp. 6) enrolled 33 HSA (19 females; age 21.91 ± 2.92 years; CSAQ-A 122.54 ± 7.34) and 31 LSA subjects (17 females; age 21.51 ± 2.33 years; CSAQ-A 71.00 ± 8.37). Both groups were matched in age and gender (Table 1).

Stimuli. Monadic walking point-light stimuli were adopted from the motion capture library [68] and processed using the Biomotion Toolbox in MATLAB [69]. Action instances start with a complete gait cycle and were clipped to be 8 seconds long. The videos featured a centrally positioned walker composed of 13 dots representing major human body joints, depicting an individual walking around in a circle (Fig 1C). Each of the angry and happy emotion categories included 45 action instances. Ambiguous monadic stimuli consisted of point-light walkers with embedded noisy dots (Fig 1C). The movement speed of the noisy dots followed the velocity distribution of human joint dots. Each noisy dot had a lifetime of 2–10 frames and moved either left or right. Their size matched that of the human joint dots.

Procedures. Point-light walkers were presented with a visual angle of 9.5° . Each biological motion stimuli were presented for 8 seconds. The other procedure and experiment settings were identical to the previous experiments.

Experiments 7–8. Clear and ambiguous dyadic biological motions. Participants. In the clear dyadic biological motion experiment (Exp. 7), we enrolled 35 HSA subjects (27 females; age 21.20 ± 2.52 years; CSAQ-A 123.20 ± 5.46)

and 35 LSA subjects (24 females; age 22.60 ± 3.70 years; CSAQ-A 67.94 ± 8.50). In the ambiguous dyadic biological motion experiment, we enrolled 35 HSA (20 females; age 21.91 ± 2.92 years; CSAQ-A 122.54 ± 7.34) and 35 LSA subjects (20 females; age 21.51 ± 2.33 years; CSAQ-A 71.00 ± 8.37). Groups were matched in age and gender (Table 1).

Stimuli and procedure. A total of 90 dyadic social interactive point-light stimuli were adopted from the Social Perception and Interaction Database [70], evenly split between angry and happy emotions (e.g., “A blamed B angrily, while B showed denial”). For each actor in dyadic interactions, the same 13 joint dots were extracted as in monadic walking stimuli (Experiment 6–7) through Autodesk Maya 2020 and were down-sampled to a frame rate of 60. Subsequently, actors were paired within each emotion category to compose dyadic biological motion stimuli (Fig 1C). Ambiguous dyadic interactive stimuli incorporate noisy dots matched in size and kinematic profile to joint dots, characterized by lifespans of 2–10 frames and consistent lateral displacement (Fig 1C). The procedure was identical to that of monadic biological motion experiments.

Behavioral analysis

Responses and RTs. Statistical analyses were conducted using R version 4.2.1 [71]. To examine whether consistent behavioral patterns emerged across experiments in relation to SA traits, we first conducted linear regression models on the ratio of negative choices and RTs in social scene tasks (Experiments 1–2), using SA traits and stimulus valence category (negative vs. positive, or neutral vs. control) as predictors. For facial expression and biological motion tasks (Experiments 3–8), we performed repeated-measures ANOVAs on accuracy and RTs, with stimulus valence (angry vs. happy) as a within-subject factor and SA group (HSA vs. LSA) as a between-subject factor.

Behavioral modeling using HDDM

Behavioral modeling was conducted using the *HDDM* package (version 0.8.0) [12]. We employed the stimulus-coding variant of the HDDM, in which the upper and lower decision boundaries correspond to the two response categories (Fig 2A). Participants’ responses and RTs were modeled using the Wiener First Passage Time (WFPT) distribution, the analytical solution of the HDDM with default diffusion noise.

The model included four free parameters (Fig 2A): initial bias (z), drift rate (v), threshold boundary (a), and non-decision time (t_0). z indicates the relative starting point between two decision boundaries, ranging from 0 to 1, where 0 corresponds to the lower decision boundary (positive responses) and 1 to the upper decision boundary (negative responses). A value of $z = 0.5$ indicates no initial bias, values exceeding 0.5 indicate initial biases toward the upper decision boundary, requiring less evidence to reach the boundary and make negative judgements. The drift rate v represents the speed and direction of signal accumulation, with a larger absolute value indicating faster and more efficient processing. The sign of v denotes the direction of accumulation, where negative values reflect evidence favoring the lower decision boundary, and positive values favor the upper decision boundary. The threshold boundary a represents the total amount of evidence required to make a decision, and greater values indicate more conservative strategies. The non-decision time t_0 captures sensory encoding and motor execution latencies.

Let $t = RT - t_0$ denote the decision time. The WFPT density for responses terminating at the boundary is [72]:

$$f(t|v, a, z) = \frac{\pi}{a^2} \exp\left(-vaz - \frac{v^2 t}{2}\right) \sum_{k=1}^{\infty} k \exp\left(-\frac{k^2 \pi^2 t}{2a^2}\right) \sin(k\pi z), \quad t > 0 \quad (1)$$

The specific HDDM model structure (i.e., parameterization of z , v , and a) was adapted to each experiment’s stimulus type, while retaining the same core HDDM framework to ensure comparability across experiments (for details see S2 Material in S1 Text). All models demonstrated satisfactory convergence ($\hat{R} \leq 1.05$) [73], and the estimated HDDM parameters were analyzed using the same statistical procedures as those applied to behavioral data.

Modeling prior bias with the social prism

To investigate how distinct social cues in the Social Prism shape prior biases in SA, we conducted a series of nested linear regression analyses using the “*statsmodels*” package (version 0.14.1) [74]. The dependent variable was the prior bias, denoted as Δz , computed as the difference between a participant’s initial bias and the LSA group mean in the same experiment (Fig 2B), as follows:

$$\Delta z_{i,e} = z_{i,e} - \frac{1}{|L_e|} \sum_{x \in L_e} z_{x,e} \quad (2)$$

where i indexes the individual participant, e denotes the experiment, and L_e refers to the set of LSA participants in experiment e .

We estimated six nested linear models of increasing complexity (Models 1–6; see S3 Material in S1 Text). Model 6, referred to as the Social Prism Model, incorporated all hypothesized social cues and their interactions with SA traits, as formulated in Eq 3:

$$\Delta z = \beta_0 + SA \cdot \left(\beta^T \mathbf{X} + AMB \cdot \gamma^T \mathbf{X}_{AMB} \right) + \varepsilon \quad (3)$$

where:

$$\beta = [\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6]^T$$

$$\gamma = [\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5]^T$$

$$\mathbf{X} = [ANI, SI, EMO, FA, BO, AMB]^T$$

$$\mathbf{X}_{AMB} = [ANI, SI, EMO, FA, BO]^T$$

In all models, social cues, including ANI, SI, EMO, FA, BO, and AMB, served as moderators of SA’s effect on prior biases. All binary-coded cues (ANI, SI, EMO, FA, BO) were categorical. AMB was treated as a continuous predictor and calculated using principles from signal detection theory (S4 Material in S1 Text, S1 Fig), reflecting the degree to which positive and negative interpretations of a social stimulus could be distinguished, with higher values indicating greater ambiguity. AMB was modeled both as a main effect and as a moderator interacting with other social cues, with the corresponding results reported in S6 Material in S1 Text. Categorical predictors were dummy-coded, and continuous predictors were used in their raw form.

Model performance was assessed by comparing the goodness-of-fit across models using the Akaike Information Criterion (AIC). Improvements in model fit were evaluated based on changes in AIC ($-\Delta AIC$) relative to a baseline intercept-only model. Nested models were compared sequentially, and a greater decrease in AIC indicated better explanatory power. This approach enabled a principled evaluation of the additional value contributed by incorporating individual social cues and their higher-order interactions with SA.

Predicting social anxiety scores with machine learning

We implemented eXtreme Gradient Boosting (XGBoost) using the “*scikit-learn*” package (version 1.7.0) [75] in Python to predict individual SA traits from behavioral and cognitive-level features. XGBoost was selected for its effectiveness in

handling multicollinearity and delivering robust performance in structured behavioral datasets [76]. The output variable was the SA traits. Three batches of input predictors were selected to examine the contributions to SA trait prediction: behavioral features only, computational features estimated via HDDM only, and a combined set comprising both. Demographic variables were included in all three clusters.

For each stimulus type, separate models were trained and evaluated using repeated hold-out (Monte Carlo cross-validation). Data were randomly split into training (80%) and testing (20%) sets across 20 independent iterations to provide a robust estimate of model performance. Within each iteration, all preprocessing and model selection were restricted to the training set. Feature and target variables were standardized on the training data and applied to the test data. XGBoost hyperparameters were optimized using Bayesian optimization, with 5-fold cross-validation on the training set as the optimization objective. Optimization proceeded for up to 100 iterations or until convergence of the mean squared error [77]. The best hyperparameter configuration was then refit on the full training set and evaluated once on the held-out test set, which was not accessed during preprocessing, tuning, or model selection. Model performance was evaluated using the average Explained Variance Score (EVS) across all cross-validation splits. EVS measured the proportion of variance in the target variable explained by the model, with values ranging from negative infinity to 1.0. Higher EVS values indicated better predictive performance. EVS scores were compared between feature sets using an independent-samples t-test within each experiment.

Social cognition stimulation with a Bayesian framework

Bayesian simulations were implemented in Python to model social cognitive processes in ambiguous social scenes (Exp. 1b), following the Bayesian inference framework. This experiment was selected because it incorporated the full set of social cues defined in the Social Prism and exhibited distinct behavioral patterns. Valence judgements were formalized as a one-step Bayesian inference process [41], in which trial-wise posterior distributions were computed by combining prior and likelihood distributions, both modeled as Gaussian (Fig 5B). A one-step framework was adopted because each ambiguous social scene was treated as an independent event, with judgments shaped by relatively stable prior expectations rather than sequential belief updating across trials.

Prior Specification. For each trial, the prior mean μ_{prior} was defined by the participant's HDDM-estimated initial bias z conditioned on the stimulus valence. Specifically, each participant i had two valence-specific initial bias parameters, $z_{\text{neg},i}$ and $z_{\text{pos},i}$, corresponding to the negative and positive conditions, respectively. These parameters were interpreted as category-specific prior expectations reflecting stable tendencies developed across repeated encounters with negative and positive social stimuli. Let $\text{val} \in \{\text{neg}, \text{pos}\}$ denote the valence of the stimulus. The valence-dependent initial bias for $z_{\text{val},i}$ was defined as:

$$z_{\text{val},i} = \begin{cases} z_{\text{neg},i}, & \text{val} = \text{neg} \\ z_{\text{pos},i}, & \text{val} = \text{pos} \end{cases} \quad (4)$$

The trial-wise prior mean was then set to this value:

$$\mu_{\text{prior}} = z_{\text{val},i} \quad (5)$$

The prior distribution was assumed to follow a normal distribution with this mean and a fixed standard deviation of $\sigma = 0.04$ to reflect moderate uncertainty:

$$\text{Prior} \sim \mathcal{N}(z_{\text{val},i}, \sigma^2) \quad (6)$$

Likelihood Estimation. To derive likelihood estimates for ambiguous social scenes, we implemented a VGG16–UMAP–KNN pipeline. This approach was intended as a computational approximation of sensory evidence rather than a biologically realistic model of human perceptual processing. Likelihoods were derived from the valence features of ambiguous social scenes (Fig 5A), estimated via a K-Nearest Neighbors (KNN) classifier trained on clear social scenes with known valence labels. Valence features were extracted using a fine-tuned VGG16 convolutional neural network [78], which achieved an accuracy of 81.82% in valence judgements for clear social scenes. Feature representations were taken from the final fully connected layer (1×4096), and subsequently reduced to a lower-dimensional space (1×5) using Uniform Manifold Approximation and Projection (UMAP) [79]. Dimensionality reduction was performed separately for clear and ambiguous images to account for distributional differences. The KNN classifier, trained on clear-scene embeddings and corresponding valence labels, was applied to ambiguous-scene embeddings to produce probabilistic valence estimates.

For each ambiguous scene stimulus s , the likelihood mean μ_s was defined based on its normalized Minkowski distances to the positive and negative cluster centroids in the 5-dimensional UMAP embedding space:

$$\mu_s = \frac{d_s^{\text{pos}}}{d_s^{\text{pos}} + d_s^{\text{neg}}} \quad (7)$$

where d_s^{pos} and d_s^{neg} denote the Minkowski distances from ambiguous stimulus s to the centroids of the positive and negative clusters, respectively. This measurement reflected the relative position of the stimulus along the positive-negative valence dimension, and was used as the mean for both types of likelihood distributions.

To quantify classification uncertainty, valence-specific standard deviations were defined as normalized deviations from each cluster center:

$$\sigma_s^{\text{pos}} = \frac{d_s^{\text{pos}}}{\bar{d}^{\text{pos}}} \quad (8)$$

$$\sigma_s^{\text{neg}} = \frac{d_s^{\text{neg}}}{\bar{d}^{\text{neg}}} \quad (9)$$

where $\bar{d}^{\text{pos}}$ and $\bar{d}^{\text{neg}}$ denoted the mean intra-cluster distances of the positive and negative training samples, respectively. To prevent the likelihood from dominating the prior due to differences in scale, we normalized all likelihood standard deviations by a scaling factor λ , set equal to the prior standard deviation ($\lambda = 0.04$), and then divided by the average unnormalized standard deviation across all stimuli. The average unnormalized standard deviation $\bar{\sigma}$ was computed as follows:

$$\bar{\sigma} = \frac{1}{2} \left(\frac{1}{N} \sum_{s=1}^N \sigma_s^{\text{pos}} + \frac{1}{N} \sum_{s=1}^N \sigma_s^{\text{neg}} \right) \quad (10)$$

where N denoted the number of ambiguous images. The normalized standard deviations for positive samples $\tilde{\sigma}_s^{\text{pos}}$ and negative samples $\tilde{\sigma}_s^{\text{neg}}$ were calculated as:

$$\tilde{\sigma}_s^{\text{pos}} = \frac{\lambda \sigma_s^{\text{pos}}}{\bar{\sigma}} \quad (11)$$

$$\tilde{\sigma}_s^{\text{neg}} = \frac{\lambda \sigma_s^{\text{neg}}}{\bar{\sigma}} \quad (12)$$

The final likelihood distribution used for trial-level simulation was defined as a Gaussian with mean μ_s , and a standard deviation determined by the valence of the presented stimulus ($\text{val} \in \{\text{neg}, \text{pos}\}$, as defined above). Specifically:

$$\text{Likelihood}_s \sim \begin{cases} \mathcal{N}(\mu_s, (\tilde{\sigma}_s^{\text{neg}})^2), & \text{val} = \text{neg}, \\ \mathcal{N}(\mu_s, (\tilde{\sigma}_s^{\text{pos}})^2), & \text{val} = \text{pos}. \end{cases} \quad (13)$$

All prior and likelihood distributions were truncated and normalized to the interval [0, 1], as both μ_{prior} and μ_s reside in a bounded decision space, ensuring compatibility in probabilistic integration.

Posterior Computation and Decision Rule. For each trial, the posterior distribution was computed by pointwise multiplication of the normalized prior and likelihood probability density functions (Fig 5B), followed by normalization:

$$\text{Posterior}(x) = \frac{\text{Prior}(x) \cdot \text{Likelihood}(x)}{\int_0^1 \text{Prior}(x) \cdot \text{Likelihood}(x) dx} \quad x \in [0, 1] \quad (14)$$

A single value x_{sample} was then drawn from the resulting posterior distribution:

$$x_{\text{sample}} \sim \text{Posterior}(x) \quad (15)$$

Based on this sampled value, the simulated binary decision $\hat{r}$ as generated using a fixed threshold at 0.5, where 0 indicates a positive judgement and 1 indicates a negative judgement:

$$\hat{r} = \begin{cases} 0, & x_{\text{sample}} < 0.5 \\ 1, & x_{\text{sample}} \geq 0.5 \end{cases} \quad (16)$$

Simulation Evaluation. Given that the simulation was parameterized using HDDM-derived priors estimated from the observed behavioral data, it was intended as a test of generative adequacy rather than an independent out-of-sample prediction. Simulation performance was evaluated in three aspects. First, for each participant, we calculated the accuracy of the model-simulated binary responses compared to the participants' observed responses, and conducted a one-sample *t*-test against 0.5 to evaluate whether the model generated random responses. Next, we assessed the correlation between the simulated and empirical negative response rates to examine the model's ability to capture overall response patterns. Third, to determine whether the simulations reflected key behavioral signatures, we computed the simulated negative response rate for each condition and participant. The simulated negative response rates were then entered into a linear regression model with SA traits and stimulus valence (positive vs. negative) as predictors, testing whether the simulated data replicated the SA-by-valence interaction observed in the empirical data.

Supporting information

S1 Text. S1 Material. Social anxiety group classification. S2 Material. HDDMs model specification. S3 Material. Linear regression models for predicting prior biases from social cues. S4 Material. Calculation of the ambiguity index. S5 Material. Additional hierarchical drift-diffusion model results. S6 Material. Ambiguity modulates cue-specific prior biases in SA. S7 Material. Machine learning predictive performance. (DOCX)

S1 Fig. Ambiguity values for ten stimulus types across experiments. Exp = Experiment. (TIF)

S2 Fig. Behavioral results and associations with SA traits. Results did not show a consistent pattern of group differences across tasks. For responses, we found a significant positive association between SA traits and the ratio of negative choices only in the negative ambiguous social scenes ($\beta = 0.238$, $R^2 = 0.057$, $p = .044$), but not for any other experiments ($p > 0.05$). For RTs, in the clear monadic biological motion experiment (Exp. 5), a repeated-measures analysis of variance

(ANOVA) revealed a significant interaction between group and stimulus valence, $F_{(1,68)} = 4.651$, $p = .035$, $\eta_p^2 = 0.064$. Post hoc comparisons showed that both the high social anxiety (HSA) group and low social anxiety (LSA) group recognized happy emotions more slowly than angry emotions, with this effect being more pronounced in the HSA group. No other significant group $\times$ stimulus valence interactions were observed in the remaining experiments. Solid lines represent linear regressions; shaded areas indicate 95% confidence intervals. Boxplots show medians and interquartile ranges; individual points reflect participants. Statistical significance: $*p < .05$, $**p < .01$, $***p < .001$ (Bonferroni-corrected). BM = biological motion; HSA = high social anxiety, LSA = low social anxiety.

(TIF)

S3 Fig. Drift rate results and associations with SA traits. Results did not show a consistent pattern of group differences across tasks. HDDM results only revealed a faster v of negative signals in the HSA group compared to LSA group in the clear monadic biological motion experiment. Solid lines represent linear regressions; shaded areas indicate 95% confidence intervals. Boxplots show medians and interquartile ranges; individual points reflect participants. Statistical significance: $*p < .05$ (Bonferroni-corrected). BM = biological motion; HSA = high social anxiety, LSA = low social anxiety.

(TIF)

S4 Fig. Decision boundary and non-decision time results and associations with SA traits. The HSA group exhibited more conservative decision-making in the neutral scenes ($\beta = -0.376$, $p = .003$) and the clear dyadic biological motions experiment ($t_{(68)} = -5.188$, $p < .001$, Cohen's $d = 1.24$). No other significant associations between SA traits and HDDM parameters were observed across experiments. Solid lines represent linear regressions; shaded areas indicate 95% confidence intervals. Boxplots show medians and interquartile ranges; individual points reflect participants. Statistical significance: $*p < .05$, $**p < .01$, $***p < .001$ (Bonferroni-corrected). BM = biological motion; HSA = high social anxiety, LSA = low social anxiety.

(TIF)

S1 Table. Machine learning predictive performance ($M \pm SD$) and statistical comparisons (t , p) between feature sets in predicting social anxiety trait scores.

(XLSX)

Acknowledgments

The authors would like to thank Zhijian Xu, Ziyao Zhang, and Jiayin Xiao for their assistance with data collection, Zhen Chen and Chenxin Wang for their help with the preparation of figures, and Zhenku Cheng for reviewing the manuscript. The authors thank Professor Roger Ratcliff (The Ohio State University) for insightful suggestions on this work.

Author contributions

Conceptualization: Yuxi Wang, Qianqian Ju, Yujia Peng.

Data curation: Yuxi Wang, Qianqian Ju, Renhe Jia, Minghao Yuan.

Formal analysis: Yuxi Wang, Qianqian Ju, Renhe Jia.

Funding acquisition: Qianqian Ju, Yujia Peng.

Investigation: Yuxi Wang, Qianqian Ju.

Methodology: Yuxi Wang, Qianqian Ju, Renhe Jia, Minghao Yuan, Yujia Peng.

Project administration: Yuxi Wang, Qianqian Ju.

Resources: Yujia Peng.

Software: Yuxi Wang, Qianqian Ju, Renhe Jia, Minghao Yuan.

Supervision: Yujia Peng.

Validation: Yuxi Wang, Qianqian Ju, Renhe Jia, Minghao Yuan.

Visualization: Yuxi Wang, Qianqian Ju, Renhe Jia, Minghao Yuan, Yujia Peng.

Writing – original draft: Yuxi Wang.

Writing – review & editing: Yuxi Wang, Qianqian Ju, Renhe Jia, Minghao Yuan, Yujia Peng.

References

1. Leung CJ, Yiend J, Lee TMC. The relationship between attention, interpretation, and memory bias during facial perception in social anxiety. *Behav Ther*. 2022;53(4):701–13. <https://doi.org/10.1016/j.beth.2022.01.011> PMID: 35697432
2. Stein MB, Stein DJ. Social anxiety disorder. *Lancet*. 2008;371(9618):1115–25. [https://doi.org/10.1016/S0140-6736\(08\)60488-2](https://doi.org/10.1016/S0140-6736(08)60488-2) PMID: 18374843
3. Heuer K, Lange W-G, Isaac L, Rinck M, Becker ES. Morphed emotional faces: emotion detection and misinterpretation in social anxiety. *J Behav Ther Exp Psychiatry*. 2010;41(4):418–25. <https://doi.org/10.1016/j.jbtep.2010.04.005> PMID: 20511123
4. Maoz K, Eldar S, Stoddard J, Pine DS, Leibenluft E, Bar-Haim Y. Angry-happy interpretations of ambiguous faces in social anxiety disorder. *Psychiatry Res*. 2016;241:122–7. <https://doi.org/10.1016/j.psychres.2016.04.100> PMID: 27173656
5. Jusyte A, Schöenberg M. Threat processing in generalized social phobia: an investigation of interpretation biases in ambiguous facial affect. *Psychiatry Res*. 2014;217(1–2):100–6. <https://doi.org/10.1016/j.psychres.2013.12.031> PMID: 24656896
6. Silvia PJ, Allan WD, Beauchamp DL, Maschauer EL, Workman JO. Biased recognition of happy facial expressions in social anxiety. *J Soc Clin Psychol*. 2006;25(6):585–602. <https://doi.org/10.1521/jscp.2006.25.6.585>
7. Goodwin I, Kugel J, Hester R, Garrido MI. Bayesian accounts of perceptual decisions in the nonclinical continuum of psychosis: greater imprecision in both top-down and bottom-up processes. *PLoS Comput Biol*. 2023;19(11):e1011670. <https://doi.org/10.1371/journal.pcbi.1011670> PMID: 37988398
8. Kok P, Mostert P, de Lange FP. Prior expectations induce prestimulus sensory templates. *Proc Natl Acad Sci U S A*. 2017;114(39):10473–8. <https://doi.org/10.1073/pnas.1705652114> PMID: 28900010
9. Aitken F, Turner G, Kok P. Prior expectations of motion direction modulate early sensory processing. *J Neurosci*. 2020;40(33):6389–97. <https://doi.org/10.1523/JNEUROSCI.0537-20.2020> PMID: 32641404
10. Dietel FA, Möllmann A, Bürkner P-C, Wilhelm S, Buhlmann U. Interpretation Bias Across Body Dysmorphic, Social Anxiety and Generalized Anxiety Disorder—A Multilevel, Diffusion Model Account. *Cogn Ther Res*. 2021;45(4):715–29. <https://doi.org/10.1007/s10608-020-10180-7>
11. Ratcliff R. A theory of memory retrieval. *Psychol Rev*. 1978;85(2):59–108. <https://doi.org/10.1037/0033-295x.85.2.59>
12. Wiecki TV, Sofer I, Frank MJ. HDDM: hierarchical bayesian estimation of the drift-diffusion model in python. *Front Neuroinform*. 2013;7:14. <https://doi.org/10.3389/fninf.2013.00014> PMID: 23935581
13. Ratcliff R, Smith PL, Brown SD, McKoon G. Diffusion decision model: current issues and history. *Trends Cogn Sci*. 2016;20(4):260–81. <https://doi.org/10.1016/j.tics.2016.01.007> PMID: 26952739
14. Mulder MJ, Prummer F, Terburg D, Kenemans JL. Drift-diffusion modeling reveals that masked faces are preconceived as unfriendly. *Sci Rep*. 2023;13(1):16982. <https://doi.org/10.1038/s41598-023-44162-y> PMID: 37813970
15. Dekel R, Sagi D, Zomet A, Levi DM, Polat U. Isolating objective and subjective filling-in using the Drift Diffusion Model. *bioRxiv*. 2023.
16. Servant M, Evans NJ. A diffusion model analysis of the effects of aging in the Flanker Task. *Psychol Aging*. 2020;35(6):831–49. <https://doi.org/10.1037/pag0000546> PMID: 32658539
17. Chen J, Short M, Kemps E. Interpretation bias in social anxiety: a systematic review and meta-analysis. *J Affect Disord*. 2020;276:1119–30. <https://doi.org/10.1016/j.jad.2020.07.121> PMID: 32777650
18. Philippot P, Douilliez C. Social phobics do not misinterpret facial expression of emotion. *Behav Res Ther*. 2005;43(5):639–52. <https://doi.org/10.1016/j.brat.2004.05.005> PMID: 15865918
19. Schneier FR, Rodebaugh TL, Blanco C, Lewin H, Liebowitz MR. Fear and avoidance of eye contact in social anxiety disorder. *Compr Psychiatry*. 2011;52(1):81–7. <https://doi.org/10.1016/j.comppsy.2010.04.006> PMID: 21220069
20. Powers KE, Worsham AL, Freeman JB, Wheatley T, Heatherton TF. Social connection modulates perceptions of animacy. *Psychol Sci*. 2014;25(10):1943–8. <https://doi.org/10.1177/0956797614547706> PMID: 25193944
21. Caballo VE, Salazar IC, Iruiria MJ, Arias B, Hofmann SG, CISO-A Research Team. The multidimensional nature and multicultural validity of a new measure of social anxiety: the Social Anxiety Questionnaire for Adults. *Behav Ther*. 2012;43(2):313–28. <https://doi.org/10.1016/j.beth.2011.07.001> PMID: 22440068
22. Wang Y, Zang Y, Peng Y. Validity and reliability of the Chinese social anxiety questionnaire for adults. *Chinese Mental Health J*. 2024;38(08):730–6.
23. Boelen PA, Reijntjes A. Intolerance of uncertainty and social anxiety. *J Anxiety Disord*. 2009;23(1):130–5. <https://doi.org/10.1016/j.janx-dis.2008.04.007> PMID: 18565725

24. Carleton RN, Collimore KC, Asmundson GJG. "It's not just the judgements--It's that I don't know": intolerance of uncertainty as a predictor of social anxiety. *J Anxiety Disord.* 2010;24(2):189–95. <https://doi.org/10.1016/j.janxdis.2009.10.007> PMID: [19931391](#)
25. Hezel DM, Stewart SE, Riemann BC, McNally RJ. Standard of proof and intolerance of uncertainty in obsessive-compulsive disorder and social anxiety disorder. *J Behav Ther Exp Psychiatry.* 2019;64:36–44. <https://doi.org/10.1016/j.jbtep.2019.02.002> PMID: [30818107](#)
26. Clark DM, Wells A. A cognitive model of social phobia. 1995.
27. Hofmann SG. Cognitive factors that maintain social anxiety disorder: a comprehensive model and its treatment implications. *Cogn Behav Ther.* 2007;36(4):193–209. <https://doi.org/10.1080/16506070701421313> PMID: [18049945](#)
28. Rapee RM, Heimberg RG. A cognitive-behavioral model of anxiety in social phobia. *Behav Res Ther.* 1997;35(8):741–56. [https://doi.org/10.1016/S0005-7967\(97\)00022-3](https://doi.org/10.1016/S0005-7967(97)00022-3) PMID: [9256517](#)
29. Beltzer ML, Adams S, Beling PA, Teachman BA. Social anxiety and dynamic social reinforcement learning in a volatile environment. *Clin Psychol Sci.* 2019;7(6):1372–88. <https://doi.org/10.1177/2167702619858425> PMID: [32864197](#)
30. Beltzer ML, Daniel KE, Daros AR, Teachman BA. Examining social reinforcement learning in social anxiety. *J Behav Ther Exp Psychiatry.* 2023;80:101810. <https://doi.org/10.1016/j.jbtep.2022.101810> PMID: [37247976](#)
31. Gibbs-Dean T, Katthagen T, Tsenkova I, Ali R, Liang X, Spencer T, et al. Belief updating in psychosis, depression and anxiety disorders: a systematic review across computational modelling approaches. *Neurosci Biobehav Rev.* 2023;147:105087. <https://doi.org/10.1016/j.neubiorev.2023.105087> PMID: [36791933](#)
32. Peng Y, Wang Y, Ju Q, Liu F, Xu J. Investigating social cognitive characteristics of social anxiety within the Bayesian framework. *Adv Psychol Sci.* 2025;33(8):1267. <https://doi.org/10.3724/sp.j.1042.2025.1267>
33. Topel S, Ma I, van Duijvenvoorde ACK, van Steenbergen H, de Bruijn ERA. Adapting to uncertainty: The role of anxiety and fear of negative evaluation in learning in social and non-social contexts. *J Affect Disord.* 2024;363:310–9. <https://doi.org/10.1016/j.jad.2024.07.066> PMID: [39043306](#)
34. Hoffmann JA, Hobbs C, Moutoussis M, Button KS. Lack of optimistic bias during social evaluation learning reflects reduced positive self-beliefs in depression and social anxiety, but via distinct mechanisms. *Sci Rep.* 2024;14(1):22471. <https://doi.org/10.1038/s41598-024-72749-6> PMID: [39341892](#)
35. Teale Sapach MJN, Carleton RN, Mulvogue MK, Weeks JW, Heimberg RG. Cognitive constructs and social anxiety disorder: beyond fearing negative evaluation. *Cogn Behav Ther.* 2015;44(1):63–73. <https://doi.org/10.1080/16506073.2014.961539> PMID: [25277488](#)
36. Anderson PL, Price M, Edwards SM, Obasaju MA, Schmertz SK, Zimand E, et al. Virtual reality exposure therapy for social anxiety disorder: a randomized controlled trial. *J Consult Clin Psychol.* 2013;81(5):751–60. <https://doi.org/10.1037/a0033559> PMID: [23796315](#)
37. Li Y, He M, Wang Z, Hofmann SG, Liu X. Distress tolerance as a mediator of mindfulness-based intervention for anxiety and depression: evidence from two randomized controlled trials. *Int J Clin Health Psychol.* 2024;24(2):100445. <https://doi.org/10.1016/j.ijchp.2024.100445> PMID: [38333547](#)
38. Li Y, Zhang Y, Wang C, Luo J, Yu Y, Feng S, et al. Supported mindfulness-based self-help intervention as an adjunctive treatment for rapid symptom change in emotional disorders: a practice-oriented multicenter randomized controlled trial. *Psychother Psychosom.* 2025;94(2):119–29. <https://doi.org/10.1159/000542937> PMID: [39809242](#)
39. Hayes SC, Luoma JB, Bond FW, Masuda A, Lillis J. Acceptance and commitment therapy: model, processes and outcomes. *Behav Res Ther.* 2006;44(1):1–25. <https://doi.org/10.1016/j.brat.2005.06.006> PMID: [16300724](#)
40. Hofmann SG, Gómez AF. Mindfulness-based interventions for anxiety and depression. *Psychiatr Clin North Am.* 2017;40(4):739–49. <https://doi.org/10.1016/j.psc.2017.08.008> PMID: [29080597](#)
41. Bitzer S, Park H, Blankenburg F, Kiebel SJ. Perceptual decision making: drift-diffusion model is equivalent to a Bayesian model. *Front Hum Neurosci.* 2014;8:102. <https://doi.org/10.3389/fnhum.2014.00102> PMID: [24616689](#)
42. Griffiths TL, Tenenbaum JB. Predicting the future as Bayesian inference: people combine prior knowledge with observations when estimating duration and extent. In: *Annual Conference of the Cognitive Science Society*, 22nd, Jul, 2000, Philadelphia, PA, US; Experiments 1–3 Were Presented at the Aforementioned Conference. Educational Publishing Foundation; 2013:56.
43. Kube T, Schwarting R, Rozenkrantz L, Glombiewski JA, Rief W. Distorted cognitive processes in major depression: a predictive processing perspective. *Biol Psychiatry.* 2020;87(5):388–98. <https://doi.org/10.1016/j.biopsych.2019.07.017> PMID: [31515055](#)
44. Palmer CJ, Lawson RP, Hohwy J. Bayesian approaches to autism: towards volatility, action, and behavior. *Psychol Bull.* 2017;143(5):521–42. <https://doi.org/10.1037/bul0000097> PMID: [28333493](#)
45. Li H, Zhao Z, Jiang S, Wu H. Brain circuits that regulate social behavior. *Mol Psychiatry.* 2025;30(7):3240–56. <https://doi.org/10.1038/s41380-025-03037-6> PMID: [40287553](#)
46. Motzkin JC, Philippi CL, Wolf RC, Baskaya MK, Koenigs M. Ventromedial prefrontal cortex is critical for the regulation of amygdala activity in humans. *Biol Psychiatry.* 2015;77(3):276–84. <https://doi.org/10.1016/j.biopsych.2014.02.014> PMID: [24673881](#)
47. Gu R, Ao X, Mo L, Zhang D. Neural correlates of negative expectancy and impaired social feedback processing in social anxiety. *Soc Cogn Affect Neurosci.* 2020;15(3):285–91. <https://doi.org/10.1093/scan/nsaa038> PMID: [32232371](#)
48. Juan Q, Shiwan T, Yurong S, Jiabo S, Yu C, Shui T, et al. Brain structural and functional abnormalities in affective network are associated with anxious depression. *BMC Psychiatry.* 2024;24(1):533. <https://doi.org/10.1186/s12888-024-05970-2> PMID: [39054442](#)

49. Brühl AB, Delsignore A, Komossa K, Weidt S. Neuroimaging in social anxiety disorder—a meta-analytic review resulting in a new neurofunctional model. *Neurosci Biobehav Rev*. 2014;47:260–80. <https://doi.org/10.1016/j.neubiorev.2014.08.003> PMID: 25124509
50. Kennedy DP, Adolphs R. The social brain in psychiatric and neurological disorders. *Trends Cogn Sci*. 2012;16(11):559–72. <https://doi.org/10.1016/j.tics.2012.09.006> PMID: 23047070
51. Qela B, Damiani S, De Santis S, Groppi F, Pichiecchio A, Asteggiano C, et al. Predictive coding in neuropsychiatric disorders: a systematic transdiagnostic review. *Neurosci Biobehav Rev*. 2025;169:106020. <https://doi.org/10.1016/j.neubiorev.2025.106020> PMID: 39828236
52. Song S, Liu A, Gao Z, Tian X, Zhu L, Shang H, et al. Event-related alpha power in early stage of facial expression processing in social anxiety: influence of language context. *Psychophysiology*. 2024;61(2):e14455. <https://doi.org/10.1111/psyp.14455> PMID: 37817450
53. Tian X, Zhu L, Zhang M, Wang S, Lu Y, Xu X, et al. Social anxiety prediction based on ERP features: a deep learning approach. *J Affect Disord*. 2024;367:545–53. <https://doi.org/10.1016/j.jad.2024.09.006> PMID: 39236887
54. Hajcak G, Foti D. Significance? Significance! Empirical, methodological, and theoretical connections between the late positive potential and P300 as neural responses to stimulus significance: an integrative review. *Psychophysiology*. 2020;57(7):e13570. <https://doi.org/10.1111/psyp.13570> PMID: 32243623
55. MacNamara A, Joyner K, Klawohn J. Event-related potential studies of emotion regulation: a review of recent progress and future directions. *Int J Psychophysiol*. 2022;176:73–88. <https://doi.org/10.1016/j.ijpsycho.2022.03.008> PMID: 35346736
56. Ollendick TH, Hirshfeld-Becker DR. The developmental psychopathology of social anxiety disorder. *Biol Psychiatry*. 2002;51(1):44–58. [https://doi.org/10.1016/s0006-3223\(01\)01305-1](https://doi.org/10.1016/s0006-3223(01)01305-1) PMID: 11801230
57. Alkozei A, Creswell C, Cooper PJ, Allen JJB. Autonomic arousal in childhood anxiety disorders: associations with state anxiety and social anxiety disorder. *J Affect Disord*. 2015;175:25–33. <https://doi.org/10.1016/j.jad.2014.11.056> PMID: 25590763
58. Mulder MJ, Wagenmakers EJ, Ratcliff R, Boekel W, Forstmann BU. Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *J Neurosci*. 2012;32(7):2335–43. <https://doi.org/10.1523/JNEUROSCI.4156-11.2012>
59. Heimberg RG. Cognitive-behavioral therapy for social anxiety disorder: current status and future directions. *Biol Psychiatry*. 2002;51(1):101–8. [https://doi.org/10.1016/s0006-3223\(01\)01183-0](https://doi.org/10.1016/s0006-3223(01)01183-0) PMID: 11801235
60. Heimberg RG, Brozovich FA, Rapee RM. A cognitive behavioral model of social anxiety disorder: update and extension. *Social Anxiety*. Elsevier; 2010. 395–422.
61. Peng Y, Knotts JD, Taylor CT, Craske MG, Stein MB, Bookheimer S, et al. Failure to identify robust latent variables of positive or negative valence processing across units of analysis. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2021;6(5):518–26. <https://doi.org/10.1016/j.bpsc.2020.12.005> PMID: 33676919
62. Peng Y, Knotts JD, Young KS, Bookheimer SY, Nusslock R, Zinbarg RE, et al. Threat neurocircuitry predicts the development of anxiety and depression symptoms in a longitudinal study. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2023;8(1):102–10. <https://doi.org/10.1016/j.bpsc.2021.12.013> PMID: 35031524
63. Faul F, Erdfelder E, Buchner A, Lang A-G. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods*. 2009;41(4):1149–60. <https://doi.org/10.3758/BRM.41.4.1149> PMID: 19897823
64. Ju Q, Xu Z, Chen Z, Fan J, Zhang H, Peng Y. Screening social anxiety with the social artificial intelligence picture system. *J Anxiety Disord*. 2025;109:102955. <https://doi.org/10.1016/j.janxdis.2024.102955> PMID: 39671733
65. Tottenham N, Tanaka JW, Leon AC, McCarry T, Nurse M, Hare TA, et al. The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Res*. 2009;168(3):242–9. <https://doi.org/10.1016/j.psychres.2008.05.006> PMID: 19564050
66. Chung K-M, Kim S, Jung WH, Kim Y. Development and validation of the yonsei face database (YFace DB). *Front Psychol*. 2019;10:2626. <https://doi.org/10.3389/fpsyg.2019.02626> PMID: 31849755
67. McKay LS, Simmons DR, McAleer P, Marjoram D, Piggot J, Pollick FE. Do distinct atypical cortical networks process biological motion information in adults with Autism Spectrum Disorders?. *Neuroimage*. 2012;59(2):1524–33. <https://doi.org/10.1016/j.neuroimage.2011.08.033> PMID: 21888982
68. Ma Y, Paterson HM, Pollick FE. A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behav Res Methods*. 2006;38(1):134–41. <https://doi.org/10.3758/bf03192758> PMID: 16817522
69. van Boxtel JJA, Lu H. A biological motion toolbox for reading, displaying, and manipulating motion capture data in research settings. *J Vis*. 2013;13(12):7. <https://doi.org/10.1167/13.12.7> PMID: 24130256
70. Okruszek Ł, Chrustowicz M. Social perception and interaction database-A novel tool to study social cognitive processes with point-light displays. *Front Psychiatry*. 2020;11:123. <https://doi.org/10.3389/fpsyg.2020.00123> PMID: 32218745
71. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing; 2022.
72. Navarro DJ, Fuss IG. Fast and accurate calculations for first-passage times in Wiener diffusion models. *J Mathematical Psychol*. 2009;53(4):222–30. <https://doi.org/10.1016/j.jmp.2009.02.003>
73. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Statist Sci*. 1992;7(4). <https://doi.org/10.1214/ss/1177011136>
74. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with python. *SciPy*. 2010;7(1):92–6. <https://doi.org/10.25080/Majors-92bf1922-011>
75. Pedregosa F, Varoquaux G, Gramfort A. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12:2825–30.

76. Chen T, Guestrin C. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016. 785–94. <https://doi.org/10.1145/2939672.2939785>
77. Nguyen V. Bayesian optimization for accelerating hyper-parameter tuning. In: 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), 2019. 302–5. <https://doi.org/10.1109/aike.2019.00060>
78. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint. 2014. <https://doi.org/arXiv:14091556>
79. McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. 2018. <https://arxiv.org/abs/1802.03426>