Face Pareidolia and the Implicit Bystander Effect: A Drift-Diffusion Model Analysis with Individual Differences in Autistic Traits

Lian H. Liang[1], Chris M. Stolle[1], Tin Yan Ng (hezulng@ln.edu.hk)[2 *], Yi Huang (yihuang@ln.edu.hk)[1,2 *]
[1]Department of Psychology, Lingnan University
[2]Wofoo Joseph Lee Consulting and Counselling Psychology Research Centre, Lingnan University
[*] corresponding authors

Abstract

The bystander effect, reduced helping in the presence of others, has been linked to automatic neural processes that inhibit action preparation. Recent work using the Drift-Diffusion Model (DDM) showed that human faces slow evidence accumulation during emergency classification tasks. We examined whether face pareidolia stimuli (objects in which illusory faces are perceived) produce similar effects, and whether autistic traits predict decision-making parameters. Participants (N = 65) classified scenes as safe or dangerous while viewing human faces, face-like objects, or non-face-like objects. Human faces significantly slowed drift rate, replicating prior findings; face pareidolia showed a similar but non-significant trend. In this non-clinical sample, Autism-Spectrum Quotient (AQ) scores predicted slower drift rates for both scene types, with modest effect sizes (r = .26 to .28), suggesting reduced excitatory efficiency with higher autistic traits. These findings extend the implicit bystander effect to individual differences and link autistic traits to evidence accumulation.

Keywords: bystander effect; drift-diffusion model; face pareidolia; autistic traits; excitation-inhibition balance

Introduction

The bystander effect describes the phenomenon where individuals are less likely to help someone in an emergency when other people are present (Darley & Latané, 1968). This effect was brought to public attention following the murder of Kitty Genovese in 1964, where numerous witnesses allegedly failed to intervene (Rosenthal, 1964). Traditional explanations for the bystander effect emphasize top-down social cognitive processes including diffusion of responsibility, evaluation apprehension, and pluralistic ignorance (Darley & Latané, 1968; Hortensius & de Gelder, 2018).

Recent neuroscience research has highlighted a complementary bottom-up mechanism. For instance, Hortensius and de Gelder (2014) used functional magnetic resonance imaging (fMRI) to examine neural responses while participants viewed emergency scenes with varying numbers of bystanders. Results showed that increased bystander presence was associated with reduced activity in motor preparation regions (precentral gyrus), while activity in mentalizing networks remained unchanged. This suggests the bystander effect operates not only through deliberate social reasoning but also via an implicit neural "brake" on automatic action readiness.

To integrate these findings, Hortensius and de Gelder (2018) proposed an opponent process model. This framework posits that witnessing emergencies triggers a competition between two motivational systems: a fast, automatic personal distress (PD) response that inhibits action through flight-or-fight mechanisms, and a slower, reflective empathic concern (EC) response that promotes altruistic action. Evidence from transcranial magnetic stimulation (TMS) studies indicates that higher PD predicts reduced motor-evoked potentials when bystanders were present, supporting the idea that bystanders amplify automatic avoidance responses (Hortensius et al., 2016).

To formally test this opponent process model, Stolle and Huang (2024) employed the Drift-Diffusion Model (DDM), a computational framework widely used to decompose reaction times into latent cognitive processes. In their paradigm, participants classified scenes as safe or dangerous while viewing smiling human faces, neutral faces, or face-colored triangles (scrambled faces, as a non-face control condition). The DDM analysis revealed that human faces significantly slowed the drift rate (the speed of evidence accumulation), operationalizing the bystander effect as reduced processing speed in the presence of face stimuli. This approach captures the implicit bystander effect: mere presence of faces—without explicit social interaction—is sufficient to alter decision-making dynamics.

However, in this study, the comparison between faces and triangles confounds social presence with visual complexity. Faces are highly salient stimuli that activate specialized neural networks, potentially slowing processing independent of any bystander-related mechanism. To address this confound, we introduced face pareidolia stimuli —objects in which observers perceive illusory faces. Recent intracranial recordings demonstrate substantial neural overlap between responses to real faces and face pareidolia across the cortical face network (Cerrahoğlu et al., 2025). We hypothesise that if the implicit bystander effect depends specifically on face-like processing rather than mere visual complexity, face pareidolia should produce intermediate slowing—stronger than non-face objects but weaker than real faces.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

Beyond stimulus effects, individual differences may modulate bystander-related processing. The DDM parameters have established mappings to neural mechanisms. Drift rate reflects excitatory efficiency via N-methyl-D-aspartate (NMDA) receptor-mediated recurrent signaling. Biophysical models demonstrate that NMDA receptors, with their slow synaptic time constant, enable temporal integration of sensory evidence over behaviorally relevant timescales; models relying solely on fast AMPA receptors fail to produce realistic decision times (Wong & Wang, 2006). In contrast, boundary separation reflects inhibitory control via gamma-aminobutyric acid (GABA)-ergic mechanisms. GABAergic interneurons provide mutual inhibition between competing neural populations, thereby determining when sufficient evidence has accumulated to overcome the inhibition and trigger a response (Bogacz et al., 2006). This computational-neural correspondence positions DDM as a behavioral window into neural excitation-inhibition (E/I) dynamics.

Autism spectrum conditions present a compelling case for examining E/I dynamics, though research has been complicated by substantial heterogeneity. The E/I imbalance hypothesis originally proposed elevated excitation in autism (Rubenstein & Merzenich, 2003). However, subsequent research reveals heterogeneous E/I patterns across individuals and brain regions, with some studies finding elevated excitation and others elevated inhibition (Dickinson et al., 2016; Sohal & Rubenstein, 2019; Trakoshis et al., 2020). This E/I heterogeneity parallels broader heterogeneity in autism: behavioral manifestations vary largely across individuals, suggesting the autism spectrum encompasses distinct subtypes (Happé & Ronald, 2008). Critically, neural heterogeneity appears to underlie this behavioral variation (Lombardo et al., 2019). Buch et al. (2023) used machine learning on functional connectivity data to identify four distinct autism subgroups, each with different connectivity patterns mapping onto distinct clinical profiles.

These findings raise a key question: if functional connectivity heterogeneity maps onto behavioral subtypes, does E/I heterogeneity show a similar structure? As a first step toward addressing this question, we examined whether the degree of autistic traits predicts DDM-indexed markers of neural E/I balance in a social perception context. This approach provides a proof of concept for linking E/I mechanisms to trait-level variation, with the expectation that future work can extend this framework to clinical symptom dimensions (e.g., social affect, repetitive behaviors) and multiple E/I indices.

The implicit bystander paradigm provides an ideal testbed for this question. Social perception difficulties—including processing faces and implied social presence—are among the most consistent findings in autism research, making this domain particularly relevant. Crucially, by examining how DDM parameters relate to autism-related trait dimensions—measured via Autism-Spectrum Quotient (AQ), Systemizing Quotient (SQ), and Empathy Quotient (EQ) —we can begin to address heterogeneity: identifying which E/I mechanism (excitation via drift rate versus inhibition via boundary separation) relates to which trait dimensions.

We tested two primary hypotheses. First, face pareidolia stimuli will produce a negative effect on drift rate similar to human faces but of lesser magnitude, consistent with partial neural overlap in face processing networks. Second, individual differences in autistic traits will predict DDM parameters, with the direction and specificity of effects informing whether particular trait dimensions are associated with altered excitation, inhibition, or both.

Methods

Participants

Participants were recruited from Prolific and received £5 upon completion of the 40-minute study, with top 10% performers receiving double incentive. From 81 initial participants, 12 who did not complete the study and 4 with accuracy below 60% were excluded following Stolle and Huang (2024), yielding a final sample of 65 participants (age range: 20-35; $M = 27.61$, $SD = 3.85$). All had normal or corrected-to-normal vision and no hand injuries. Ethical approval for this study was granted in accordance with Lingnan University's research ethics procedures, and informed consent was obtained from all participants.

Materials and Procedure

The experiment was administered online using lab.js. Participants were randomly assigned to one of three between-subjects conditions: real-face, face-like object (pareidolia), or non-face-like object.

Prior to the main task, participants completed three 10-item questionnaires: the Autism-Spectrum Quotient (AQ-10; Allison et al., 2012), Systemizing Quotient-Revised (SQ-R; Wheelwright et al., 2006), and Empathy Quotient (EQ; Baron-Cohen & Wheelwright, 2004).

In the scene classification task, participants judged whether scenes were "dangerous" or "safe" by pressing J or K (mapping reversed between two blocks of 180 trials each). Each trial presented a stimulus image at the top of the screen for 2-4 seconds, followed by a scene image requiring classification within 1.5 seconds (Figure 1).

Scene images (384 total; 182 per category: safe and dangerous) were sourced from Stolle and Huang (2024). Real-face stimuli were sourced from the same study; face-like objects were selected from the ECCV "FaceInThings" dataset (Hamilton et al., 2024); non-face-like objects were identical images with facial features removed via Photoshop. The number of stimuli (0, 3, or 6) varied randomly across trials (Figure 2).

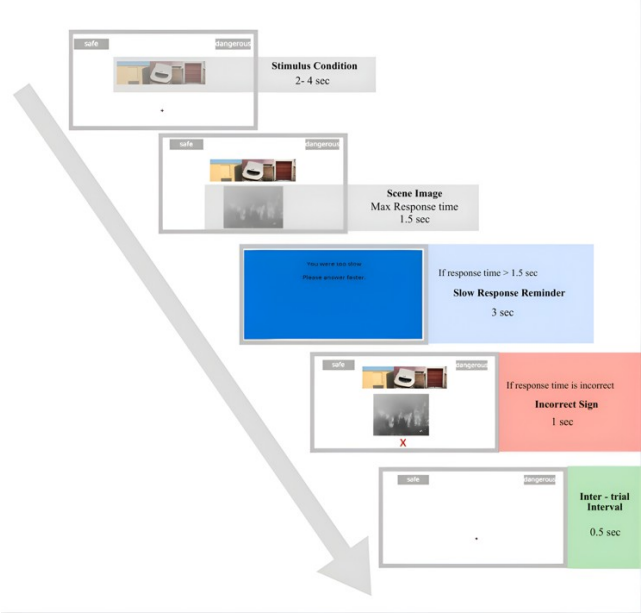

Figure 1: Example of trial showing possible feedback. BLUE: time-out. RED: incorrect response.

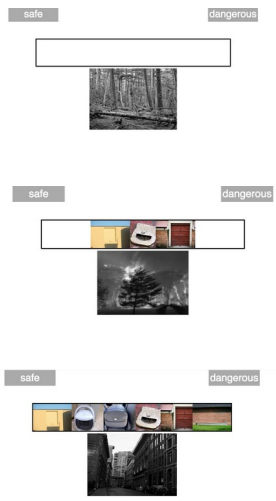

Figure 2: Example trials with 0 , 3 and 6 non-face-like object image.

DDM Parameter Estimation

The DDM decomposes reaction times into four parameters: bias (β, starting point of evidence accumulation), boundary separation (α, amount of evidence required for decision), drift rate (v, speed of evidence accumulation), and non-decision time (τ, encoding and motor components). Parameters were estimated using a hierarchical Bayesian approach, allowing group-level effects while retaining individual-level estimates. Specifically, we fitted a Wiener distribution to predict reaction times in seconds, with dangerous decisions coded positively (+RT) and safe decisions negatively (-RT). Individual parameters for bias (β), threshold (α), and non-decision time (τ) were drawn from normal distributions with sample-wide means (Mu) and standard deviations. We also estimated two independent drift rates (v): one for safe (negative) scenes and one for dangerous (positive) scenes, both drawn hierarchically from sample-wide normal distributions. Additionally, we included a bystander effect parameter for subjects in the human-face condition (added to the safe-condition drift rate) and a separate parameter for the face-like object condition (subtracted from the risk-condition drift rate), each drawn from its own group-level normal prior. This asymmetric specification reflects the opponent-process logic of the paradigm (Hortensius & de Gelder, 2018). Human face stimuli plausibly bias decisions toward 'safe' by signalling the presence of a supportive or non-threatening social agent, so their effect was modelled as an additive shift on the safe-decision drift. By contrast, a face-like object is expected to reduce the perceived authenticity or threat value of a dangerous scene by introducing a benign agentic cue, which was modelled as a subtractive shift on the risk-decision drift. Estimating the two parameters independently allows each condition's effect to be captured on its own theoretically motivated dimension rather than imposing a single symmetric slowdown.

Model estimation was conducted through R in Jags (Su & Yajima, 2021) using the Wiener distribution implementation (Wabersich & Vandekerckhove, 2014). Six MCMC chains each ran 5,000 iterations with a burnin of 2,000 iterations and a thinning rate of 4. Diagnostics confirmed all R-hat < 1.01 and effective sample size (ESS) > 1,000 for all parameters.

Analysis Plan

We excluded fast guesses with reaction time (RT) < 300ms and missing answers following Stolle and Huang (2024). Note, the maximum RT allowed was 1.5 seconds.

For behavioral analyses, we conducted Welch's t-tests comparing RTs and two-sample proportion tests comparing accuracy across conditions. Given our theoretical predictions were directional and pairwise – comparing human faces and face-like objects independently to the non-face control – we used targeted t-tests and condition-specific DDM bystander parameters rather than an omnibus one-way ANOVA, which could obscure the specific face-vs-object and pareidolia-vs-object contrasts of interest.

For DDM analyses, we examined posterior means, probability of direction (PD), and highest density intervals (HDI). For individual differences, we computed Pearson correlations between questionnaire scores and DDM parameters.

Results

Behavioral Results

Human-face vs. Non-face-like object. RTs in the human-face condition (M = 528.37, SD = 87.28) were significantly slower than in the non-face-like condition (M = 524.40, SD

= 82.58), $t(11479) = -2.79$, $p = .01$, $d = -.06$. This difference was significant for safe scenes ($p = .02$) but not dangerous scenes ($p = .10$). Accuracy was significantly lower in the human-face condition (88.61%) compared to non-face-like (90.69%), $\chi^2 = 16.98$, $p < .001$.

Face-like object vs. Non-face-like object. RTs in the face-like condition ($M = 519.13$, $SD = 89.53$) differed from non-face-like ($M = 524.40$, $SD = 82.58$), $t(4730) = -2.79$, $p < .05$, $d = -.06$, though the direction indicated faster responses for face-like stimuli. This pattern was significant for dangerous scenes ($p < .05$) but not safe scenes ($p = .13$). Accuracy did not differ between conditions ($p = .89$).

DDM Results

The results of the DDM parameters estimation are listed in Table 1. We found a threshold α of 1.242. There is a slight bias β toward safe decisions, with both the mean and the HDI being below the neutral value of 0.5. The non-decision time is .326, indicating that participants needed at least 326ms time for responding which does not account for their decision-making. The drift rate has a slightly larger value for dangerous decisions compared to safe decisions. However, note that in combination with the bias towards safe decisions, this may be explained through the additional required evidence-accumulation to overcome the bias.

Table 1: DDM Parameter Estimates

DDM Parameters Estimation	Posterior Mean	Probability of Direction (PD)	Highest Density Interval (HDI)
αMu (Boundary)	1.242	--	[1.201, 1.288]
βMu (Bias)	0.485	--	[0.474, 0.496]
τMu (Non-decision)	0.326	--	[0.319, 0.332]
vDanger (Drift)	2.853	--	[2.628, 3.097]
vSafe (Drift)	-2.564	--	[-2.796, -2.347]
Bystander Effect (Human)	-0.314	0.980*	[-0.628, -0.017]
Bystander Effect (Face-like)	-0.153	0.836	[-0.453, 0.178]

Note. Probability of direction. .97* .99** .999***

Human-face stimuli produced a significant negative effect on drift rate (posterior mean = -0.314, PD = 0.980, HDI = [-0.628, -0.017]), indicating slower evidence accumulation. Face-like stimuli showed a similar negative trend (posterior mean = -0.153, PD = 0.836) but the HDI included zero ([-0.453, 0.178]), indicating the effect was not statistically reliable.

Questionnaire-DDM Correlations

Drift rate associations. AQ was significantly correlated with the drift rate for both scene types. For safe scenes, higher AQ was associated with slower drift, $r = .26$, $t(61) = 2.08$, $p = .04$. For dangerous scenes, this effect was stronger, $r = -.38$, $t(61) = -3.25$, $p = .002$. Note that positive correlation with drift_safe and negative correlation with drift_danger both indicate slower evidence accumulation toward the respective decision boundary for individuals with higher AQ. Neither SQ (rs = -.05 to -.10, ps > .41) nor EQ (rs = -.02 to -.08, ps > .54) showed significant associations with drift rate.

Boundary separation associations. No questionnaire scores significantly predicted boundary separation (α): AQ, $r = -.10$, $p = .44$; SQ, $r = -.09$, $p = .47$; EQ, $r = -.01$, $p = .92$.

Discussion

This study investigated the implicit bystander effect using face pareidolia stimuli and examined individual differences in autistic traits as predictors of decision-making parameters. Human faces reliably slowed drift rate during the scene classification task, replicating the implicit bystander effect within a DDM framework. Face pareidolia stimuli showed a numerically similar but statistically inconclusive trend. Higher autistic trait levels predicted slower drift rate but not boundary separation, implicating reduced excitatory efficiency rather than generalised E/I dysregulation. Together, these findings partially support the face-specificity hypothesis and provide novel evidence that DDM-indexed E/I markers track individual differences in autistic traits.

At the behavioral level, human faces slowed responses and reduced accuracy, consistent with the bystander effect; the unexpected faster responses for face-like objects likely reflect their lower social salience. The DDM analysis clarifies these patterns: replicating Stolle and Huang (2024), human faces significantly slowed drift rate during scene classification, consistent with the opponent process model whereby face presence activates automatic avoidance responses that impede decision-making. Face pareidolia stimuli showed a similar negative trend but did not reach statistical significance. This pattern aligns with neural evidence of overlapping but non-identical activation for real and illusory faces (Cerrahoğlu et al., 2025). Real faces may trigger additional social-cognitive processes—including semantic interpretation, emotion processing, and identity recognition—that amplify the bystander effect beyond what pareidolia stimuli evoke. The apparent discrepancy between faster behavioural RTs and the negative drift rate trend for

face-like objects is consistent with the decomposition the DDM affords: total RT reflects the joint contribution of drift rate, non-decision time, and boundary separation (Ratcliff & McKoon, 2008). Face-like objects may reduce non-decision time or boundary separation through heightened perceptual salience speeding stimulus encoding, while simultaneously slowing evidence accumulation process. The behavioural t-test collapses across these components, whereas the DDM isolates drift rate as the index most directly tied to the opponent-process mechanism of interest. Future research with larger samples and controlled stimulus sets may clarify whether face pareidolia produces reliable, if attenuated, bystander effects.

The individual differences findings provide preliminary evidence linking autistic trait levels to DDM-indexed markers of neural E/I balance in social perception. Higher AQ scores predicted slower drift rates for both safe and dangerous scene classifications—suggesting reduced excitatory efficiency that manifests consistently across scene types within this social perception context. Critically, AQ did not predict boundary separation, indicating that higher autistic trait levels were associated with reduced excitation but not altered inhibition—a pattern consistent with a lower E/I ratio. This dissociation suggests that autism-related social processing differences may be characterized by selectively reduced excitatory efficiency rather than generalized E/I dysregulation.

As a proof of concept, these findings offer preliminary support for the broader hypothesis that E/I heterogeneity may map onto variation in autistic traits, paralleling Buch et al.'s (2023) demonstration that functional connectivity heterogeneity relates to distinct autism subtypes. That AQ specifically predicted drift rate—while SQ and EQ did not, raises the possibility that reduced E/I ratio may characterize certain autism-related dimensions more than others. This preliminary dissociation aligns with fractionation models proposing partially independent etiologies for different trait dimensions (Happé & Ronald, 2008), though replication with comprehensive behavioral profiling is needed. Establishing robust trait-mechanism correspondences will require examining multiple trait dimensions across diverse E/I indices.

The absence of SQ and EQ associations is consistent with this fractionated view. That empathizing tendencies (EQ) did not predict DDM parameters is notable given that the bystander paradigm involves social-emotional face processing—a domain where empathy-related differences might be expected. This null finding suggests that the implicit bystander effect and its modulation by autistic traits operate through mechanisms distinct from explicit empathic processing, reinforcing the need for future studies to systematically map which E/I markers relate to which trait dimensions.

Several limitations warrant consideration and point toward future directions. First, the face pareidolia stimuli were not matched for perceived "faceness" across participants; individual differences in pareidolia perception may have attenuated effects. Second, stimulus sets differed in color properties, introducing potential confounds. Third, we measured autistic traits using the AQ-10, which captures a single broad dimension. More comprehensive behavioral profiling—covering social communication, repetitive behaviors, and sensory processing dimensions that show dissociable neural correlates (Buch et al., 2023)—would enable systematic testing of which trait dimensions map onto which E/I profiles. Future analyses will decompose AQ into Austin's (2005) sub-scales to test whether specific trait dimensions differentially drive drift rate effects, given that particular autistic trait dimensions may be more relevant to specific task demands than others. Fourth, drift rate provides one index of neural E/I balance; convergent validation using neurophysiological measures such as magnetic resonance spectroscopy or EEG-based E/I ratio indices would strengthen these interpretations. Fifth, our cross-sectional design cannot address whether E/I-related processing differences are stable trait characteristics or state-dependent. These limitations notwithstanding, the current findings establish a proof of concept for linking DDM-indexed E/I markers to individual differences in autistic traits.

Conclusion

This study extends the implicit bystander effect framework by demonstrating that autistic traits predict evidence accumulation efficiency indexed by DDM. By revealing that AQ specifically—but not SQ or EQ—predicts drift rate, these findings provide initial evidence that E/I mechanisms may differentially relate to distinct trait dimensions. This proof-of-concept approach—examining which neural mechanisms relate to which traits in which contexts—offers a framework for future studies to systematically map E/I heterogeneity onto the autism subtype structure, potentially informing more precise characterization of autism-related processing.

Acknowledgments

The work described in this paper was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. LU 23600225), by the General Research Fund of Research Grants Council in Hong Kong to Yi Huang (13610724) and by the Wofoo Joseph Lee Consulting and Counselling Psychology Research Centre, Lingnan University.

Declaration of Generative AI use

During the preparation of this work the authors used GPT and DeepSeek in order to polish the writing. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). Toward brief "Red Flags" for autism screening: The Short

Autism Spectrum Quotient and the Short Quantitative Checklist for Autism in toddlers in 1,000 cases and 3,000 controls [corrected]. *Journal of the American Academy of Child and Adolescent Psychiatry*, *51*(2), 202–212.e7. https://doi.org/10.1016/j.jaac.2011.11.003

Austin, E. J. (2005). Personality correlates of the broader autism phenotype as assessed by the Autism Spectrum Quotient (AQ). *Personality and Individual Differences*, *38*(2), 451–460. https://doi.org/10.1016/j.paid.2004.04.022

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of autism and developmental disorders*, *34*(2), 163–175. https://doi.org/10.1023/b:jadd.0000022607.19833.00

Buch, A. M., Vértes, P. E., Seidlitz, J., Kim, S. H., Grosenick, L., & Liston, C. (2023). Molecular and network-level mechanisms explaining individual differences in autism spectrum disorder. *Nature neuroscience*, *26*(4), 650–663. https://doi.org/10.1038/s41593-023-01259-x

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review*, *113*(4), 700–765.

Cerrahoğlu, B., Jacques, C., Rekow, D., Jonas, J., Colnat-Coulbois, S., Caharel, S., Leleu, A., & Rossion, B. (2025). The neural basis of face pareidolia with human intracerebral recordings. *Imaging neuroscience (Cambridge, Mass.)*, *3*, imag_a_00518.

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: diffusion of responsibility. *Journal of personality and social psychology*, *8*(4), 377–383. https://doi.org/10.1037/h0025589

Dickinson, A., Jones, M., & Milne, E. (2016). Measuring neural excitation and inhibition in autism: Different approaches, different findings and different interpretations. *Brain research*, *1648*(Pt A), 277–289. https://doi.org/10.1016/j.brainres.2016.07.011

Hamilton, M. *et al.* (2025). Seeing Faces in Things: A Model and Dataset for Pareidolia. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G. (eds) Computer Vision – ECCV 2024. ECCV 2024. Lecture Notes in Computer Science, vol 15123. Springer, Cham. https://doi.org/10.1007/978-3-031-73650-6_22

Happé, F., & Ronald, A. (2008). The 'fractionable autism triad': a review of evidence from behavioural, genetic, cognitive and neural research. *Neuropsychology review*, *18*(4), 287–304. https://doi.org/10.1007/s11065-008-9076-8

Hortensius, R., & de Gelder, B. (2014). The neural basis of the bystander effect--the influence of group size on neural activity when witnessing an emergency. *NeuroImage*, *93 Pt 1*, 53–58. https://doi.org/10.1016/j.neuroimage.2014.02.025

Hortensius, R., & de Gelder, B. (2018). From Empathy to Apathy: The Bystander Effect Revisited. *Current directions in psychological science*, *27*(4), 249–256. https://doi.org/10.1177/0963721417749653

Hortensius, R., Schutter, D. J., & de Gelder, B. (2016). Personal distress and the influence of bystanders on responding to an emergency. *Cognitive, affective & behavioral neuroscience*, *16*(4), 672–688.

Lombardo, M. V., Lai, M. C., & Baron-Cohen, S. (2019). Big data approaches to decomposing heterogeneity across the autism spectrum. *Molecular psychiatry*, *24*(10), 1435–1450. https://doi.org/10.1038/s41380-018-0321-0

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Rosenthal, A. M. (1964). *Thirty-eight witnesses: The Kitty Genovese case*. McGraw-Hill.

Rubenstein, J. L., & Merzenich, M. M. (2003). Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes, brain, and behavior*, *2*(5), 255–267. https://doi.org/10.1034/j.1601-183x.2003.00037.x

Sohal, V. S., & Rubenstein, J. L. R. (2019). Excitation-inhibition balance as a framework for investigating mechanisms in neuropsychiatric disorders. *Molecular psychiatry*, *24*(9), 1248–1257. https://doi.org/10.1038/s41380-019-0426-0

Stolle, C. M., & Huang, Y. (2024). The Mere Reminder of Others: A Cognitive Modelling Approach to the Implicit Bystander Effect. In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the CognitiveScience Society* (Vol. 46, pp. 5766-5772) https://escholarship.org/uc/item/6p71c47t

Su, Y.-S., & Yajima, M. (2021). R2jags: Using r to run 'jags' [Computer software manual]. (R package version 0.7-1)

Trakoshis, S., Martínez-Cañada, P., Rocchi, F., Canella, C., You, W., Chakrabarti, B., Ruigrok, A. N., Bullmore, E. T., Suckling, J., Markicevic, M., Zerbi, V., MRC AIMS Consortium, Baron-Cohen, S., Gozzi, A., Lai, M. C., Panzeri, S., & Lombardo, M. V. (2020). Intrinsic excitation-inhibition imbalance affects medial prefrontal cortex differently in autistic men versus women. *eLife*, *9*, e55684. https://doi.org/10.7554/eLife.55684

Wabersich, D., & Vandekerckhove, J. (2014). Extending JAGS: A tutorial on adding custom distributions to JAGS (with a diffusion model example). *Behavior research methods*, *46*(1), 15-28.

Wheelwright, S., Baron-Cohen, S., Goldenfeld, N., Delaney, J., Fine, D., Smith, R., Weil, L., & Wakabayashi, A. (2006). Predicting Autism Spectrum Quotient (AQ) from the Systemizing Quotient-Revised (SQ-R) and Empathy Quotient (EQ). *Brain research*, *1079*(1), 47–56. https://doi.org/10.1016/j.brainres.2006.01.012

Wong, K. F., & Wang, X. J. (2006). A recurrent network mechanism of time integration in perceptual

decisions. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *26*(4), 1314–1328.
https://doi.org/10.1523/JNEUROSCI.3733-05.2006