Substantial heritability underlies fairness norm adaptation capability and its neural basis

Authors: Yuening Jin, Dang Zheng, Ruolei Gu, Qingchen Fan, Martin Dietz, Changshuo Wang, Xinying Li, Jie Chen, Yuanyuan Hu, Yuan Zhou, Yuan Zhou

Date: 2024-09-09T00:00:00+00:00

Abstract

本研究揭示了公平规范适应能力、其神经相关物以及长期心理健康结果的共同遗传基础。研究招募了 186 对双胞胎，在其成年早期作为回应者参与最后通牒博弈（UG）的同时接受功能性磁共振成像（fMRI）扫描（研究一），并在八年后测量其抑郁症状（研究二）。通过计算建模，UG 中的规范适应过程与公平估值过程得以区分。这两个过程均具有中等程度的遗传力。前脑岛表现出显著表型相关，而辅助运动区/内侧额回（SMA/mSFG）则与规范适应指标——学习率——表现出显著表型相关及共同的遗传影响。多巴胺能 DRD2 多态性与学习率以及 SMA/mSFG 对预测误差的编码均存在相关，构成了它们的共同遗传基础。进一步的区域基因表达分析揭示了 SMA/mSFG 中多巴胺相关基因的高表达。此外，学习率可预测八年后的抑郁症状严重程度，DRD2 多态性构成了二者的共同遗传基础。这表明遗传力是规范适应背后不可忽视的驱动力，它促进变化环境中社会规范的学习并维持长期心理健康。

Full Text

Preamble

Substantial heritability underlies fairness norm adaptation capability and its neural basis

Yuening Jin[1,2], Dang Zheng[1,3], Ruolei Gu[1,2], Qingchen Fan[1,2], Martin Dietz[4], Changshuo Wang[1,5,6], Xinying Li[7,2], Jie Cheng[7,2], Yuanyuan Hu[7,2], Yuan Zhou[1,2,8]

[1]CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China
[2]Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China
[3]Department of Early Childhood Education, China National Children's Center, Beijing 100035, China
[4]Center of Functionally Integrative Neuroscience, Institute of Clinical Medicine, Aarhus University, Universitetsbyen 3, 8000 Aarhus C, Denmark

[5]Sino-Danish Center, University of Chinese Academy of Sciences, Beijing 100049, China
[6]Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[7]CAS Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China
[8]The National Clinical Research Center for Mental Disorders & Beijing Key Laboratory of Mental Disorders, Beijing Anding Hospital, Capital Medical University, Beijing, China

Corresponding author: Yuan Zhou, Ph.D., Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Chaoyang District, Beijing 100101, P.R. China; Email: zhouyuan@psych.ac.cn

Competing Interest Statement: All authors have approved the current version of the manuscript and its submission. All authors report no biomedical financial interests or potential conflicts of interest.

Author Contribution Statement: YJ: conceptualization, methodology, software, formal analysis, validation, writing – original draft, visualization. DZ: conceptualization, methodology, software, formal analysis, investigation, data curation, writing – review & editing. RG: conceptualization, writing – review & editing, supervision. QF: conceptualization, visualization, writing – review & editing. MD: conceptualization, writing – review & editing, supervision. CW: conceptualization, methodology, formal analysis, visualization, writing – review & editing, supervision. XL: conceptualization, data curation, writing – review & editing, supervision. JC: conceptualization, data curation, writing – review & editing, supervision. YH: conceptualization, data curation, writing – review & editing. YZ: conceptualization, methodology, software, validation, resources, writing – review & editing, supervision, project administration, funding acquisition.

Ethics Approval Statement: The study was approved by the Institutional Review Board of the Institute of Psychology, Chinese Academy of Sciences.

Funding Sources: This work was supported by the National Natural Science Foundation of China (Nos. 81771473, 82171535, 72033006), STI2030-Major Projects (2021ZD0200600), and the Technical Support Talents Project of Chinese Academy of Sciences (grant number: E2CX1154).

Data Availability: Data and code that support the findings of this study will be available upon reasonable request to the corresponding author.

Abstract

The present research uncovers the shared genetic underpinnings of fairness norm adaptation capability, its neural correlates, and long-term mental health outcomes. We recruited 186 twins who played as responders in the Ultimatum Game (UG) while undergoing fMRI scanning in early adulthood (Study-1) and

measured their depressive symptoms eight years later (Study-2). Using computational modeling, we differentiated the process of norm adaptation from fairness valuation in the UG. Both processes exhibited moderate heritability. The anterior insula showed a significant phenotypic correlation, whereas the supplementary motor area/medial superior frontal gyrus (SMA/mSFG) demonstrated both a significant phenotypic correlation and shared genetic influence with the learning rate—an index of norm adaptation. Dopaminergic DRD2 polymorphisms correlated with both the learning rate and SMA/mSFG encoding of prediction error, constituting their common genetic basis. Further regional gene expression analysis revealed high expression of dopamine-related genes in the SMA/mSFG. Moreover, the learning rate predicted depressive symptom severity eight years later, with DRD2 polymorphisms constituting their shared genetic basis. These findings suggest that heritability is a non-negligible driving force behind norm adaptation, which facilitates learning of social norms in changing environments and preserves long-term mental health.

Keywords: fairness norm adaptation; prediction error encoding; twin study; genetics; anterior insula; supplementary motor area

1. Introduction

Cultural Evolution Theory posits that stability and change represent two major mechanisms through which social norms evolve. While adaptive norms endure across generations, maladaptive norms are supplanted by new norms through social learning processes to ensure viability in altered social conditions. The field has already recognized the role of heritability in fueling one dynamic of norm evolution—namely, norm stability—implying that norms crucial for survival are readily transmitted across generations. This phenomenon expedited our evolution into a modern society that inherently upholds social norms of reciprocity and obligation. Could heritability also fuel the other dynamic of norm evolution: norm change? Norm change essentially relies on rapid acquisition of adaptive norms. Despite its profound survival implications, the heritability of norm acquisition capability remains scarcely investigated. The rapid acquisition of adaptive norms (i.e., high social learning capability) transmitted across generations could grant us an evolutionary advantage similar to the transmission of adaptive norms themselves. Only when heritability fuels both sides of norm dynamics can we both quickly develop and preserve adaptive norms, thereby acquiring a survival advantage in changing environments.

Fairness norms, being among the most prevalent types of social norms, exhibit significant variability across situations. A strong capability to learn what constitutes an acceptable fair offer in different contexts from different stakeholders holds profound survival implications. On one hand, in non-challenging times, enforcing norms against malicious transgressors safeguards our self-interests, preserves our dignity, and deters potential transgressors. On the other hand, in challenging times, timely adjustment of norm standards—setting minimal expectations on others—becomes crucial, allowing us to both dissipate negative

feelings and accumulate resources offered by stringent allocators. A strong adaptation capability thus grants both resources and psychological wellbeing in the long run. Multiple lines of empirical evidence suggest that heritability constitutes an indispensable force in shaping the capability to learn fairness norms. First, domain-general cognitive aspects of social norm learning processes (such as cognitive flexibility) show more than 30% heritability. Second, domain-specific processes of moral appraisal associated with fairness norms, along with their relevant neural activities, also exhibit more than 30% heritability. Conversely, environmental influences play only a limited role in shaping social learning capabilities. However, it remains to be determined whether the heritability of our ability to learn social norms within the fairness domain and its underlying neural basis can be established.

The lack of studies investigating the genetic contribution to fairness norm learning and adaptation can be attributed, in part, to technical challenges in quantification. However, recent developments in computational modeling have addressed these challenges, allowing for differentiation of a norm adaptation process from a fairness valuation process in fairness decision-making during the UG. The system of norm adaptation (indexed by the learning rate), which captures how quickly one learns and adjusts one's norm, is both conceptually and neurologically distinct from the system of fairness valuation—indexed by the initial fairness norm and fairness sensitivity—which captures norm strictness and sensitivity. We are interested in whether the norm adaptation system exhibits a distinct pattern of genetic influence different from the fairness valuation system.

Following heritability analyses, this study hypothesized that genetic expression related to neurotransmitter activity, particularly the dopamine D2 receptor gene (DRD2), which strongly modulates dopaminergic levels, would be associated with norm adaptation ability and its underlying neural activities. Previous studies have documented dopaminergic influence on both fairness-specific and general learning processes. Enhanced dopamine levels increased sensitivity to both advantageous and disadvantageous inequity proposals in the Dictator Game, reinforcing the representation of social norms in people's minds and prompting them to reduce inequality. Additionally, studies have found that enhanced dopamine levels accelerate the learning process by facilitating minimization of prediction errors (PE)—i.e., narrowing the gap between actual rewards and expectations. Previous studies have identified that multiple DRD2 polymorphisms, linking with D2 receptor density, were related to individual capabilities in avoiding negative outcomes in probabilistic learning tasks. Genotypes associated with higher D2 receptor availability (i.e., the A1 allele carrier in rs1800497 and the T/T homozygotes of rs6277) have consistently shown superior performance in learning to avoid negative consequences across studies. Based on these findings, we propose that dopaminergic genes modulating dopamine levels could affect the rate of norm adaptation. We focused on three SNPs that strongly influence D2 receptor density: rs1800497, rs2283265, and rs6277. The prevalence of the A1 allele (T) in both the DRD2/ANKK1-Taq Ia (rs1800497) and rs2283265 polymorphisms leads to a 30% reduction in D2 receptor density compared to

the A2 allele (C-G), resulting in lower dopamine levels. The prevalence of T in DRD2 C957T (rs6277) leads to enhanced density of striatal D2 receptors. We thus hypothesized that the T allele in DRD2 genotypes (rs1800497 and rs2283265) and the G allele in rs6277 would reduce the learning rate for fairness norms.

While the association between dopaminergic neurons and reinforcement signals is well-established, the role of serotonin in learning processes has been less clear. Some evidence suggests that serotonin is involved in controlling impulsivity—i.e., inhibition of an impulsive response upon viewing an unfair offer. Recent studies have found that serotonin is specifically associated with learning to avoid negative events but does not directly pertain to norm learning. We thus hypothesized that only dopaminergic, but not serotonergic SNPs, would influence the learning rate for norms—in other words, a predominant dopaminergic genetic influence on the norm adaptation system.

On the other hand, we hypothesized that the fairness valuation system may be simultaneously modulated by dopaminergic and serotonergic genetic influences. Previous studies have documented modulation of both serotonergic and dopaminergic systems on the proposal allocation process in the DG/UG, which involves detection of fairness violation. Serotonin level has consistently been negatively associated with rejection rates toward unfair offers in the UG, as observed through experimental manipulations and natural observations in multiple studies. It is possible that serotonergic SNPs, which modulate serotonin level, would influence the fairness valuation system denoted by the strictness of fairness norm and fairness sensitivity in the UG. Among serotonergic SNPs, the most extensively studied influencer of social behavior is the tryptophan hydroxylase-2 gene (TPH2, rs4570625). Previous studies have found that the T allele in rs4570625 generally relates to increased serotonin levels and promotes social cooperation. The influence of rs4570625 extends to prediction of trait emotional instability and occurrence of major depressive disorder. Since fairness decisions in the UG are emotionally demanding, we focus on exploring whether TPH2 would influence fairness valuation in the UG. Specifically, we hypothesized that the T allele in the TPH2 genotype would be associated with a less strict (lower) initial fairness norm imposed on proposers and lower fairness sensitivity. The role of the dopaminergic system in modulating fairness decisions has also been well-documented. Higher dopamine levels are generally associated with willingness to seek higher personal economic incomes rather than to altruistically punish others. Accordingly, we hypothesized that the T allele in rs1800497 and rs2283265, as well as the G allele in rs6277—which relate to reduced dopamine levels—would be associated with higher (stricter) initial fairness norms and higher fairness sensitivity.

Taking a further step, we investigated the heritability and dopaminergic genetic modulation of the neural basis of norm adaptation. Learning of fairness norms essentially involves encoding the difference between the expected norm and the newly-encountered split ratio—i.e., the encoding of norm PE. Previous studies

have shown convergent involvement of the insula and a range of brain regions related to PE encoding, including the medial orbitofrontal cortex (mOFC), ventromedial prefrontal cortex (vmPFC), and substantia nigra/ventral tegmental area (SN/VTA). To date, the genetic influences on the neural encoding of norm PE remain unknown.

Taking another step further, we hope to discover both the phenotypic and common genetic connections between norm adaptation capability, its neural basis, and adaptation consequences for long-term mental health, to provide preliminary support for our hypothesis that a genetically-powered adaptation capability would grant us a biological advantage for long-term survival. A longitudinal design spanning eight years gives us an opportunity to answer this question.

In brief, the current study endeavors to investigate the genetic basis underlying norm adaptation ability and its neural activities with a relatively large twin sample through a series of analyses across two studies (Fig. 1). In Study-1 (fMRI experimental study), participants performed a modified version of the UG while undergoing functional magnetic resonance imaging (fMRI) (Fig. 2). We used a computational modeling approach to separate the adaptation system from the fairness valuation system in the UG and delved into their heritability. Further, we examined whether unique dopaminergic genetic influences are responsible for the norm adaptation system, which could be distinct from the genetic underpinnings of the fairness valuation system. We then used univariate genetic modeling and bivariate genetic modeling analyses to further identify brain regions whose activity, modulated by PE, not only demonstrates phenotypic correlation but also shares common genetic influences with the learning rate. Last, we explored whether dopamine-related gene activity constitutes the shared genetic foundation, with a particular focus on candidate D2 genetic polymorphisms. Therefore, complementary to the SNP analyses, we provided additional supporting evidence for broader dopaminergic expression in brain regions responsible for encoding norm PE signals, drawing from the Allen Human Brain Atlas (AHBA). In Study-2 (longitudinal survey), we recalled the twin cohort eight years after Study-1 and traced their depression symptoms. We investigated whether norm adaptation capability could predict the occurrence of depressive symptoms in the long term and whether these traits shared a common dopaminergic genetic basis.

2.1 GLMM

(Please insert Fig. 1 about here) (Please insert Fig. 2 about here)

We used generalized mixed effects modeling (GLMM) to investigate the main and interaction effects of proposer type, split ratio, and time on the binomial decision outcome of acceptance or rejection in the UG. We found significant main effects of split ratio ($z = 34.36$, $p < .001$), proposer type ($z = 8.01$, $p < .001$), and trial number ($z = 3.03$, $p = .003$). Participants had significantly lower acceptance rates toward human proposers than computer proposers (esti-

mate (SE) = -.60 (.07), z = -8.13, p < .001). We found a significant interaction between split ratio and proposer type (z = -2.22, p = .027). The effect of split ratio was significantly higher in the human than computer proposer condition (estimate (SE) = 1.36 (.62), z = 2.22, p = .027), suggesting higher fairness sensitivity toward human than computer proposers and motivating the specification of two separate fairness sensitivity parameters for human and computer proposers in the computational model.

We also found a significant trial number by proposer type interaction (z = -2.72, p = .006). Participants showed a larger time effect when facing human proposers than computer proposers (estimate (SE) = .015 (.005), z = 2.78, p = .006). Specifically, the time effect was only significant for human proposers (estimate (SE) = .011 (.004), 95%CI = [.004, .019]) and not for computer proposers (estimate (SE) = -.003(.004), 95%CI = [-.011, .004]). This suggested that acceptance rate increases with time when participants face only human proposers, even after controlling for ratio effects, implying the presence of a norm adaptation process specific to the human proposer condition.

2.2 Computational Modeling

Among various candidate models depicting the decision process in the UG, model comparison favored Model-1, which had the lowest LOOIC (Supplementary Material Table S2-3). Model-1 assumed a learning process for only human proposers but not computer proposers, consistent with our GLMM results. The 95% highest density intervals (HDIs) for group-level parameters are shown in Table S2-4. Parameter estimation revealed higher fairness sensitivity toward human than computer proposers, echoing model-free GLMM analyses (95% HDI: [.02, .05]). Model recovery suggested that simulated responses from Model-1 had a predictive accuracy of 91.21% on the original response and could capture key characteristics of the original response (Supplementary Material S6). Parameter recovery suggested a high correlation between original and recovered parameters (Supplementary Material S7).

In the optimal model (Model-1), we found a high Pearson correlation between the initial fairness norm and fairness sensitivity (r = .87), implying high internal consistency between the two indices of the fairness valuation process. We also found a low Pearson correlation (r = -.18) between the initial fairness norm and the learning rate, implying that the fairness valuation process (indexed by the initial fairness norm and fairness sensitivity) was phenotypically distinct from the learning process (indexed by the learning rate). In the next section, we investigate the heritability underlying these two distinct processes.

2.3 Heritability Analysis on Behavioral Indicators

The intraclass correlations (ICCs) for parameters of Model-1 and acceptance rates toward human and computer proposers are shown in Table 1 . Comparison of ICCs using Fisher's r-to-z transformation suggested (marginally) significantly

higher ICCs for monozygotic (MZ) than dizygotic (DZ) twins for the initial fairness norm, learning rate, fairness sensitivity toward human proposers, and acceptance rate toward human proposers. We conducted univariate heritability analyses exclusively on these parameters.

The AE model provided the optimal fit for all variables, indicated by its lowest AIC compared to alternative models (Table 1). Genetic contributions accounted for 37% of variance in acceptance rate toward human proposers. Additionally, genetic factors contributed to 41% of variance in the initial fairness norm, 36% of variance in the learning rate, and 32% of variance in fairness sensitivity toward human proposers. (Please insert Table 1 about here)

2.4 Voxel-wise Univariate Genetic Modeling on PE Maps

The AE model had the lowest AIC for most voxels (47,819 out of 47,930 voxels), so we specified the AE model for analyses. Only brain regions showing a genetic effect with $\geq 90\%$ posterior confidence in the posterior prediction map (PPM) were considered to have significant additive genetic influence. Consequently, 18 brain regions were identified from the PPM of PE as regions of interest (ROIs) for subsequent analyses (Fig. 3), which included the medial prefrontal cortex, supplementary motor area (SMA)/medial superior frontal gyrus (mSFG), right anterior insula, lateral prefrontal cortices, temporal regions, posterior parietal cortex, visual cortices, caudate, parahippocampus, and cerebellum. The mean level of genetic effect across voxels for each ROI is shown in Supplementary Material S8 Table S8-1. In the next section, we further test whether their encoding of PE shows phenotypic correlation and shares common genetic influences with the learning rate for each ROI. (Please insert Fig. 3 about here)

2.5 Bivariate Genetic Modeling

The AE model had the lowest AIC in 15 out of 18 bivariate genetic models (Supplementary Material S8 Table S8-2). In cases where AE models were not optimal, neither AE nor the optimal CE model showed significant differences from the full ACE model. We thus chose the AE model for subsequent bivariate genetic modeling analyses. The phenotypic correlation and common genetic influence between each ROI and model parameters are shown in Supplementary Material S8 Table S8-3.

The parametric estimate value for PE in the SMA/mSFG had a significant phenotypic correlation (r = -.34, 95% CI = [-.47, -.20], $p < .0001$) and a significant genetic correlation (r_g = -.61, 95% CI = [-1.00, -.61], $p < .0001$) with the learning rate. The parametric estimate value for PE in the anterior insula also showed a phenotypic correlation with the learning rate (r = -.27, 95% CI = [-.41, -.12], $p = .0002$) but did not show significant shared genetic influence with the learning rate (r_g = -.36, 95% CI = [-.80, .23], $p = .172$). These results indicated that higher learning rates were associated with larger magnitude encoding of negative PE in both the SMA/mSFG and anterior insula (Fig. 4A

[Figure 4: see original paper] and 4B), implying that individuals with higher learning capability had more sensitive neural representation of how actual split ratios violated the fairness norm. (Please insert Fig. 4 about here)

2.6 Candidate Gene-Brain-Behavior Association Analysis

The frequency distribution of SNPs in the current study showed no significant differences from the distribution of SNPs in the East Asian population in the 1000 Genomes Project (Supplementary Material S9). We found that dopaminergic SNPs influenced the learning rate and brain activities modulated by PE encoding. Hierarchical Linear Models (HLMs) revealed that rs1800497, rs2283265, and their additive score significantly (or marginally) correlated with both the learning rate and the parametric estimate for PE in the SMA/mSFG. However, rs6277 was neither predictive of the learning rate nor the neural activities in the SMA/mSFG (Table 2). We found a unique dopaminergic influence on the neural encoding of PE, as evidenced by the effect of dopaminergic additive score on both the learning rate and SMA/mSFG activity, even after controlling for TPH2 (Table 4). This finding suggested that dopaminergic genes, rather than TPH2, constituted the shared genetic influence affecting both the learning rate and encoding of PE in neural activities (Table 2).

We further found a significant mediation effect of SMA/mSFG activity on the relationship between (1) rs1800497 and the learning rate (Sobel test statistics = 1.96, p = .050) (Fig. 5A), (2) rs2283265 and the learning rate (Sobel test statistics = 2.23, p = .026) (Fig. 5B), and (3) dopaminergic additive score and the learning rate (Sobel test statistics = 2.13, p = .033) (Fig. 5C). These results implied that dopaminergic SNPs modulated norm adaptation capability via SMA/mSFG encoding of PE. (Please insert Fig. 5 about here) (Please insert Table 2 about here)

Additionally, we found that distinct genetic contributions underlay the norm adaptation versus fairness valuation systems. HLM revealed that the contributions of rs1800497, rs2283265, and rs4570625 to the initial fairness norm were significant or marginally significant both when entered separately and together (Table 3). Serotonergic and dopaminergic SNPs contributed a portion of unique variance to the initial fairness norm (Table 3). We derived a similar finding for the other indicator of the fairness valuation system, namely fairness sensitivity (Supplementary Material S10). (Please insert Table 3 about here)

2.7 Gene Enrichment Analysis on Genes Prominently Expressed in the SMA/mSFG

The top 5% expressed genes (n = 785) located in the SMA/mSFG region were identified as key genes prominently expressed in this region. Gene ontology (GO) biological processes and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways related to the key gene list were aligned using the Metascape online toolbox. After correcting for enrichment terms (pFDR < .05)

and removing discrete enrichment clusters, the top-20 significant GO biological processes and KEGG pathways included dopaminergic enrichment synapses (hsa04728), modulation of chemical synaptic transmission (GO: 0050804), and vesicle-mediated transport in synapse (GO: 0099003) (Fig. 6A [Figure 6: see original paper]). This indicated that chemical synaptic transmission, especially dopaminergic pathways and relevant biological processes, were highly active in SMA/mSFG, potentially playing a key role in this region's functionality. The Metascape enrichment network visualization (Fig. 6B) further revealed that dopaminergic synapse occupied the hub spot of the network, emphasizing again the important role that dopamine-related gene activity plays in the SMA/mSFG. (Please insert Fig. 6 about here)

2.8 Predicting Long-term Depressive Symptoms

Learning rate was predictive of depressive symptoms measured by the Beck Depression Inventory (BDI-II) ($F(1,118) = 4.67$, $p = .033$) eight years later. Bivariate genetic modeling within the optimal AE model further revealed that the two variables shared a significant phenotypic correlation ($r = -.22$, 95% CI = [-.40, -.02], $p = .023$) and a marginally significant common genetic basis ($r_9 = -.41$, 95% CI = [-.80, .01], $p = .050$), with depressive symptoms showing a high level of heritability (67%, 95%CI = [36%, 83%]). Moreover, we found a marginally significant mediation effect of learning rate between DRD2 gene polymorphisms and depressive symptoms (Sobel test statistics = -1.36, $p = .087$). That is, DRD2 gene polymorphisms could enhance norm adaptation capability (estimate (SE) = 6.86e-4 (3.94e-4), $p = .084$), which in turn buffered against the occurrence of depressive symptoms eight years later (estimate (SE) = -351.72 (161.055), $p = .031$).

3. Discussion

The current research conducted a series of analyses across two studies to investigate the heritability of norm adaptation ability, its underlying neural mechanisms, and its association with long-term mental health outcomes. We also explored the DRD2 and serotonergic genetic polymorphisms underlying these processes. In Study-1 (fMRI experimental study), we found moderate heritability underlying both the overall acceptance rate and sub-processes extracted by computational modeling, including: (1) the norm adaptation process, denoted by the learning rate, and (2) the fairness valuation process, denoted by norm strictness (initial fairness norm) and fairness sensitivity. Only dopaminergic polymorphisms (rs1800497, rs2283265, and their additive scores) influenced the learning rate, which could be distinct from the serotonergic and dopaminergic genetic underpinnings of the fairness valuation system. Among the 18 identified regions modulated by PE and showing credible genetic influence via the univariate genetic modeling approach, we further found that the anterior insula had a significant phenotypic correlation but non-significant shared genetic correlation with the learning rate, whereas the SMA/mSFG showed both significant phe-

notypic correlation and shared genetic correlation with the learning rate. The dopaminergic DRD2 polymorphisms simultaneously correlated with the learning rate and SMA/mSFG encoding of norm PE. Furthermore, SMA/mSFG encoding of norm PE mediated the relationship between DRD2 polymorphisms and the learning rate. Genes associated with dopaminergic enrichment synapses (hsa04728), modulation of chemical synaptic transmission (GO: 0050804), and vesicle-mediated transport in synapse (GO: 0099003) were highly expressed in the SMA/mSFG, providing further evidence for dopamine-related gene activity in this region. In Study-2 (longitudinal survey), we found that norm adaptation capability and long-term mental health (i.e., self-reported depression) shared both phenotypic correlation and common genetic influences, and that DRD2 influenced mental health via the mediating role of norm adaptation capability.

Our study elucidated, for the first time, moderate heritability underlying two crucial and distinct processes involved in fairness decisions: the fairness valuation process and the norm adaptation process. In this study, the initial fairness norm and fairness sensitivity represented a similar underlying construct in the fairness valuation process, displaying high correlation between them (r = .87). Conversely, the initial fairness norm appeared phenotypically distinct from the learning rate, as indicated by a low Pearson correlation (r = -.18), echoing findings by Gu et al. The fairness valuation process was subject to influence from both dopaminergic and serotonergic SNPs, whereas the norm adaptation process was solely influenced by dopaminergic SNPs, consistent with previous observations of dopaminergic modulation of learning. These results provide further support for the distinct nature of these two processes.

Furthermore, we found that individuals with varying learning rates showed differences in the strength of norm PE encoding in the anterior insula. That is, individuals with faster norm learning showed larger magnitude negative parametric correlations of anterior insula activation with norm PE, implying that individuals with faster learning rates have stronger insula activation related to encoding negative PEs. Previous studies have documented the role of insula in representing error signals along various norm dimensions and initiating actions or belief modification to diminish such errors. Insula lesions result in failures to initiate actions that adjust subjective norms. The right anterior insula serves to detect norm deviations in the UG and Trust Game, consistent with the lateralization of our major findings. Previous studies have also documented insula activation upon receiving unfair treatments in the UG and upon experiencing aversive emotions such as sadness, anger, and disgust. Our study further suggests that the anterior insula also encodes norm PE signals and that individuals with varying norm adaptation capabilities differ in neural encoding of norm PE in this region.

The common genetic influence of norm PE encoding in the anterior insula with the learning rate did not reach significance. This might imply that different genetic bases underlie learning capability and insula activity. Additionally, our supplementary analyses found non-significant influence of dopaminergic SNPs

on insula activity (Supplementary Material S10). This negative finding may be accounted for by functional heterogeneity of insula in emotional appraisal and unfairness representation, which makes it potentially susceptible to influence from a wider heterogeneous variety of SNPs beyond those dopaminergic SNPs modulating the learning rate. Future studies could conduct SNP genotyping on a wider range of genes to search for SNPs that differentially influence the learning rate and insula activity.

Our findings revealed a phenotypic correlation between the strength of norm PE encoding in the SMA/mSFG and the learning rate. This aligns with Chang and Sanfey's finding that encoding of norm PE is associated with the SMA, and with SMA's function in unfairness appraisal and rejection behavior in the UG. Our finding is also consistent with SMA's role in learning, including value computation of options in non-social probabilistic learning tasks and action monitoring and error detection in Simon tasks.

Bivariate genetic modeling analysis further revealed a common dopaminergic genetic influence between the learning rate and norm PE encoding in the SMA/mSFG. Furthermore, DRD2 influences learning ability via the mediating role of SMA/mSFG encoding of norm PE. Results from gene enrichment analysis of SMA/mSFG provide additional support for high dopamine-related gene activity in this area. Adding to dopaminergic modulation in fronto-striatal encoding of PE signals, our study provides new evidence that the SMA/mSFG is also subject to dopamine modulation in PE encoding.

Our study, for the first time, established a robust phenotypic and genetic correlation between norm adaptation capability and long-term adaptive consequences on mental health, illustrating that our dopaminergic-genetically powered learning capability could indeed grant us a survival advantage in terms of buffering against depression. DRD2 variants have been identified as biomarkers for major depressive disorders and predictors of social withdrawal. Our study revealed a new mediating path of social adaptation capability through which DRD2 influences mental health.

This study has several limitations. First, we examined the influence of a limited number of dopaminergic and serotonergic SNPs on the learning rate and initial fairness norm, as well as the neural correlates of norm learning. We also utilized a relatively simple method to estimate additive genetic effects—calculating the additive score of two DRD2 SNPs (rs1800497 and rs2283265). Although we found complementary evidence about dopamine-related gene activity in SMA/mSFG via gene enrichment analysis, future research could conduct genotyping on a wider variety of SNPs or perform genome-wide association studies to explore additive influences of a broader range of genes on norm learning capability and its neural correlates. Second, although our sample size is comparable to other twin studies employing fMRI data, it remains limited compared to behavioral studies (without fMRI) on the heritability of social norms and fairness appraisal. Future studies could validate the heritability of norm adaptation capability with behavioral experiments using larger sample sizes.

In conclusion, our study reveals that heritability is a non-negligible driving force behind learning of social norms and its neural basis, which enhances rapid acquisition of adaptive values across generations, granting us an evolutionary advantage in changing environments in terms of maintaining long-term mental health. As norm adaptation constitutes one important aspect of norm evolution dynamics, these findings, for the first time, elucidate the critical role of heritability in fueling social norm evolution and mental health. Since dopamine-related gene activity underlies norm adaptation and its neural basis, our study provides potential dopaminergic genetic markers for identifying individuals at higher risk for social maladaptations and depressive symptoms. Furthermore, our study revealed that anterior insula and SMA/mSFG encoding of norm PE may serve as heritable neural signatures of norm adaptation capability. Future studies could further explore the role of these two regions in encoding violations of a broader range of social norms. To elucidate the causal relationship between dopamine level and norm adaptation capabilities, future studies could directly manipulate dopamine levels to observe corresponding changes in fairness adaptation and its neural activities.

5.1 Participants

To ensure adequate sample size for studying brain-behavior relationships, in Study-1 we initially recruited 100 pairs of same-sex twins from the Beijing Twins Brain-Behavior Association Project, which was dependent on the Beijing Twin Study. Eight participants and their twin siblings in seven pairs (two from the same pair) were excluded due to not understanding instructions or providing random behavioral responses. This resulted in valid data from 93 pairs of twins (52% female; 48 monozygotic pairs (MZ), 45 dizygotic pairs (DZ)) for behavioral analyses. Their age ranged from 16 to 26 years (M = 20.27; SD = 2.36). Six additional pairs of twins were excluded due to excessive head motion (see fMRI data acquisition and preprocessing subsection), resulting in 87 pairs of twins (54% female; 44 MZ, 43 DZ; age M = 20.21; SD = 2.24) in the fMRI analyses. In Study-2, we successfully recalled 122 participants eight years later in December 2023 and administered the BDI-II, a self-report measure of depressive symptoms. None of the participants had self-reported current/history of physical/psychiatric diagnoses, neurological or metabolic illnesses, or head injuries. All participants read and signed informed consent before the experiment.

5.2 Task and Experimental Design of UG

Participants played a one-shot anonymous UG in the E-prime 2.0 environment while undergoing functional magnetic resonance imaging (fMRI). Scanning comprised two sessions of equal length with a 20-second break between them. All participants played the role of UG responder in all trials (Fig. 2 [Figure 2: see original paper]). Participants received a financial reward as a token for participation plus remuneration determined by their actual income from two randomly selected trials. Their choices of acceptance or rejection would influence pro-

posers' remuneration as well, which would be distributed to the corresponding players at the experiment's conclusion.

Participants played 48 trials total: 24 with human proposers and 24 with computer proposers. A full list of proposal splits is shown in Supplementary Material S1-1. We fixed the amount of money offered to the responder at one of 9, 10, or 11 yuan. We manipulated fairness level by varying the ratio distributed to the responder against the sum of money distributed to the proposer and responder. We included 24 ratio types ranging from 4% to 50%. This study adopted a within-subject design of 2 (proposer type) $\times$ 24 (ratio).

5.3 Analytical Strategies Overview

The analytical procedures comprised behavioral and neuroimaging analyses (Fig. 1 [Figure 1: see original paper]). In behavioral analyses, we first conducted a model-free generalized linear mixed-effects model (GLMM) to inspect key features of behavioral responses, which provided insights for computational model specification. We then built candidate models, performed model selection, model recovery, and parameter recovery. Within the optimal model, we conducted heritability analyses on acceptance rates and model parameter estimates using a univariate genetic modeling approach. In neuroimaging analyses, we conducted a first-level GLM for each participant and calculated the PE contrast map for the human proposer condition for each individual. We then performed voxel-wise univariate genetic modeling and identified voxels with $\geq$ 90% posterior probability of significant genetic influence in the posterior probability map (PPM). Extracting clusters with size $\geq$ 20 from the PPM, we estimated phenotypic correlation and common genetic influence between the mean activation of these clusters in the PE contrast map and the learning rate using bivariate genetic modeling. We further identified common dopaminergic SNPs underlying brain and behavior when they shared common genetic influences. We also investigated the phenotypic correlation and common genetic basis between learning capabilities and depressive symptoms eight years later.

5.4 Study-1 (fMRI Experimental Study)

Generalized Linear Mixed Effects Model (GLMM)

We first conducted model-free analyses to clarify (1) whether participants had different fairness sensitivity toward human versus computer proposers and (2) whether acceptance rate evolved with time when participants faced human versus computer proposers. If yes for (1), fairness sensitivity should be estimated separately for human and computer proposers in computational models. If yes for (2), time should influence decision outcomes after controlling for proposer type and ratio, implying a changing fairness norm. We used a logistic GLMM to investigate main and interaction effects of proposer type, split ratio, and time on the binomial decision outcome of acceptance or rejection with the lme4 and

emmeans packages in R 4.1.3. We estimated random intercepts for family and subject.

Computational Modeling

Following previous studies, we specified three critical processes for all candidate norm adaptation models: a value calculation process, a decision process, and a norm adaptation process.

Value calculation process. Individuals calculate the utility of acceptance and rejection options to make a decision. Similar to Gu et al., the utility of acceptance reflected monetary utility and Fehr-Schmidt inequality aversion utility:

$$U(\text{accept}) = (1-\beta)\cdot r(\text{responder}) + \beta \cdot r(\text{responder}) \cdot \left[\frac{r(\text{proposer}) - r(\text{responder})}{r(\text{proposer}) + r(\text{responder})} - \text{fairness\_norm} \right]$$

Inequality aversion involves comparing the encountered split ratio with the fairness norm in mind—an internal representation of what constitutes fairness. This comparison was transformed to the same scale as monetary utility by multiplying by the amount of money received by the responder, then weighted by a free parameter: fairness sensitivity. We also weighted the bare monetary gain to (1) keep monetary utility and inequality aversion utility on the same scale and (2) ensure that higher fairness sensitivity corresponded to lower preference for money. We specified different fairness sensitivity for human and computer proposers (β_h and β_c) if participants showed different fairness sensitivity toward human versus computer proposers in the GLMM analysis. The utility of rejection was 0, as both players gained nothing:

$$U(\text{reject}) = 0$$

Decision process. We used a softmax function to link option utility with individuals' probability of choosing that option:

$$p(\text{accept}) = \xi \cdot \frac{\exp(\tau \cdot U(\text{accept}))}{\exp(\tau \cdot U(\text{accept})) + \exp(\tau \cdot U(\text{reject}))} + \frac{1-\xi}{2}$$

We included two free parameters: τ depicts the level of randomness in choice, and ξ depicts occasional unintended choices by mistake. Both represent noise in decisions.

Candidate norm adaptation processes. We built multiple candidate models to represent the norm adaptation process. Previous studies consistently show that participants have an internal representation of the fairness norm—the fairness level expected from a virtual human proposer on each trial—and can flexibly adjust the norm according to interaction histories to adapt to changing environments in single-shot UG. Therefore, we assumed that fairness norm learning

occurs when participants interact with human proposers. As previous modeling studies did not include computer proposers, we assumed that when facing computer proposers, participants either had no learning process:

$$\text{fairness\_norm}_c(t) = \text{fairness\_norm}_c(t-1)$$

or learned at the same rate as human proposers where fairness norms toward human versus computer proposers update separately:

$$\text{fairness\_norm}_h(t) = \text{fairness\_norm}_h(t-1) + \alpha_h \cdot \left[\frac{r(\text{proposer})}{r(\text{proposer}) + r(\text{responder})} - \text{fairness\_norm}_h(t-1) \right]$$

$$\text{fairness\_norm}_c(t) = \text{fairness\_norm}_c(t-1) + \alpha_c \cdot \left[\frac{r(\text{proposer})}{r(\text{proposer}) + r(\text{responder})} - \text{fairness\_norm}_c(t-1) \right]$$

update together:

$$\text{fairness\_norm}(t) = \text{fairness\_norm}(t-1) + \alpha \cdot \left[\frac{r(\text{proposer})}{r(\text{proposer}) + r(\text{responder})} - \text{fairness\_norm}(t-1) \right]$$

learned at a different rate compared with human proposers, or other variations. The formulas represent how individuals adjust their subjective fairness norm based on prediction error—the difference between the current trial's split ratio and the pre-existing norm in mind—weighted by the learning rate α, a free parameter.

We specified the initial fairness norm at the beginning of the UG as a free parameter. In candidate models, we either specified the same or different initial fairness norms for human and computer proposers, generating 8 candidate models (Supplementary Material Table S2-1).

We used the Hierarchical Bayesian estimator in Rstan to simultaneously derive group-level and individual estimates for all participants. The ranges and prior distributions for each parameter at group and individual levels are shown in Supplementary Material Table S2-2. We compared goodness of fit using the leave-one-out information criterion (LOOIC), which calculates pointwise out-of-sample prediction accuracy using the log-likelihood of simulated posterior parameter values. We performed model recovery to test whether simulated data from the winning model could capture key characteristics in the original data. We performed parameter recovery to examine the robustness of parameter estimation (for details, see Supplementary Material S3).

Heritability Analyses of Behavioral Indicators

To determine heritability underlying behavioral indicators, we first calculated intraclass correlation (ICC) and performed genetic modeling analyses on rejection rate in the UG and key parameters (i.e., initial fairness norm, learning rate, and fairness sensitivity) in the winning computational model, controlling for age and sex. We then performed genetic modeling analysis with the OpenMx package in R 3.1.2 to partition additive genetic (A) (referred to as heritability), shared environmental (C), and non-shared environmental (E) contributions to variance in behavioral indicators. We fitted both the full ACE model and various sub-models (i.e., AE, CE, and E), selecting the optimal model based on chi-square changes, Akaike Information Criterion (AIC), and the principle of parsimony.

fMRI Data Acquisition and Preprocessing

Magnetic resonance images were acquired on a 3 Tesla GE Discovery MR750 MRI scanner at the Institute of Psychology, Chinese Academy of Sciences, Beijing. Image acquisition details are presented in Supplementary Material S4.

First-level GLM

As norm adaptation entails computation of prediction errors (PE), we conducted first-level general linear model (GLM) analyses to discover brain regions that encode PE (split ratio in the current trial minus fairness norm in the previous trial) for each individual. Specifically, we used parametric analysis to identify brain regions modulated by PE. Each GLM included three main events: proposal display, decision screen, and outcome display in a one-factorial (proposer type) design matrix constructed by convolving each event onset with a canonical hemodynamic response function. The PE for each trial extracted from the computational model was entered into the GLM as a parametric regressor at the proposal display event. Residual effects of head motion were accounted for by including estimated six motion parameters for each subject as covariates. We then built the PE contrast map for each individual at the proposal revelation event in the human proposer condition, including only the human proposer condition because both GLMM and the optimal model revealed that learning occurred only toward human proposers, not computer proposers.

Voxel-wise Univariate Genetic Modeling on PE Maps

We performed voxel-wise univariate genetic modeling to estimate genetic, shared environmental, and non-shared environmental contributions to prediction error (PE) t-maps. We fitted both the full ACE model and various sub-models (i.e., AE, CE, and E), selecting the optimal model based on chi-square changes, AIC, and the principle of parsimony. For almost all voxels, the best-fitting model was AE (see Results section). We then determined the significance of A by comparing chi-square with and without A (AE model vs. E model). A significant ($p < 0.05$) decrease in chi-square indicates a significant contribution of A on each voxel. We then constructed a posterior probability map (PPM) for PE to identify regions with $\geq 90\%$ posterior confidence showing credible additive genetic effect. PPM marks the posterior probability that an effect exceeds

a particular threshold (a prior mean of zero in SPM) and enables Bayesian inference about regionally specified effects, thus having no multiple comparison issues. We extracted all clusters with voxel sizes greater than 20 for subsequent analyses as our ROIs. We then calculated average heritability in each ROI and averaged parametric estimate values in each ROI for each participant for subsequent analyses.

Bivariate Genetic Modeling

We performed bivariate genetic modeling for the learning rate and average parametric estimate in each ROI. This analysis identifies brain regions that exhibit significant phenotypic correlation with the learning rate (Bonferroni correction $p < 0.05$/number of ROIs) while examining common genetic influence between model parameters and brain activity. Specifically, we decomposed the phenotypic correlation into ACE components using a correlated factors model and compared it with various sub-models (AE, CE, and E). We selected the optimal model based on chi-square changes, AIC, and the parsimony principle. The significance of bivariate genetic correlation was determined by whether the 95% CI contains 0.

Candidate Genes–Brain–Behavior Association Analysis

In the current study, we focused on DRD2/ANKK1-Taq Ia (rs1800497), DRD2 C957T (rs6277), rs2283265, and TPH2 (rs4570625) polymorphisms. We conducted SNP genotyping for all participants (see Supplementary Material S5 for procedures). We examined whether our SNP distribution followed Hardy-Weinberg Equilibrium by comparing with SNP distribution in large East Asian populations from the 1000 Genomes Project. We further used Hierarchical Linear Models (HLMs) (estimating random intercept of family) to identify dopaminergic and serotonergic SNPs simultaneously related to the learning rate and brain activity modulated by PE. For SNPs that significantly predicted both the learning rate and brain activation, we further tested for the mediating role of brain activity between SNPs and the learning rate.

Regional Gene Expression in SMA/mSFG

The AHBA dataset provided brain-wide gene expression data with 3,702 spatially distinct samples collected from six postmortem brains. Preprocessing followed previously published guidelines: (1) each probe was annotated to genes using the Re-annotator toolbox; (2) probes were filtered, excluding probes not exceeding background signal in 50% of all samples across six subjects; (3) probe selection, choosing the probe with highest differential stability across six subjects as the representative probe for each gene; (4) considering limited samples in the right hemisphere, samples were mirrored to the opposite hemisphere and samples with mean distance to ROI less than 2mm were assigned to the target region (SMA/mSFG); (5) for each sample, gene expression was normalized using scaled robust sigmoid transformation; (6) genes were filtered based on differential stability across six subjects. The whole process was implemented with

the Abagen toolbox. Here, SMA/mSFG was the ROI and mean gene expression was calculated as the regional expression level for each gene.

The top 5% (n = 785) highly expressed genes were identified as prominently expressed genes.

Enrichment Analysis

For each given gene list, pathway and process enrichment analyses were carried out by Metascape analysis (https://metascape.org/gp/index.html#/main/step1), with the following ontology sources: KEGG Pathway, GO Biological Processes, Reactome Gene Sets, Canonical Pathways, CORUM, WikiPathways, and PANTHER Pathway. All genes in the genome were used as the enrichment background. Terms with p-value < 0.01, minimum count of 3, and enrichment factor > 1.5 (the ratio between observed counts and counts expected by chance) were collected and grouped into clusters based on membership similarities. Specifically, p-values were calculated based on the cumulative hypergeometric distribution, and q-values were calculated using the Benjamini-Hochberg procedure to account for multiple correction. Kappa scores were used as the similarity metric when performing hierarchical clustering on enriched terms, and sub-trees with similarity > 0.3 were considered a cluster. The most statistically significant term within a cluster was chosen to represent the cluster.

To further capture relationships between terms, a subset of enriched terms was selected and rendered as a network plot, where terms with similarity > 0.3 were connected by edges. We selected terms with the best p-values from each of the 20 clusters, with constraints of no more than 15 terms per cluster and no more than 250 terms total. The network was visualized using Cytoscape, where each node represented an enriched term and was colored by its cluster ID (Fig. 6B).

5.5 Study-2 (Longitudinal Survey)

Predicting Long-term Depressive Symptoms

We used HLM to examine the effect of learning rate on BDI-II scores eight years later, controlling for random slope of family. We performed bivariate genetic modeling to examine whether learning rate and depressive symptoms shared phenotypic correlation and common genetic correlation. We also conducted mediation analysis to examine whether DRD2 polymorphisms would affect depressive symptoms via the mediating effect of learning rate.

Acknowledgement

We thank Jie Zhang and Ting Chen from the Institute of Psychology, Chinese Academy of Sciences for assistance with data collection in Study-1. We thank Ting Chen and Qingwen Ding from the Institute of Psychology, Chinese Academy of Sciences for assistance with data collection in Study-2.

References

1. Boyd, R. & Richerson, P.J. *Culture and the evolutionary process* (University of Chicago Press, 1988).
2. Henrich, J. *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter* (Princeton University Press, 2016).
3. Kandler, C., Penner, A., Richter, J. & Zapko-Willmes, A. The Study of Personality Architecture and Dynamics (SPeADy): A longitudinal and extended twin family study. *Twin Research and Human Genetics* 22, 548-553 (2019).
4. Zakharin, M. & Bates, T.C. Testing heritability of moral foundations: Common pathway models support strong heritability for the five moral foundations. *European Journal of Personality* 37, 485-497 (2023).
5. Graham, J., Haidt, J. & Nosek, B.A. Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology* 96, 1029 (2009).
6. Haidt, J. The new synthesis in moral psychology. *Science* 316, 998-1002 (2007).
7. Hertz, U. Learning how to behave: Cognitive learning processes account for asymmetries in adaptation to social norms. *Proceedings of the Royal Society B* 288, 20210293 (2021).
8. Kölle, F. & Quercia, S. The influence of empirical and normative expectations on cooperation. *Journal of Economic Behavior & Organization* 190, 691-703 (2021).
9. Vavra, P., Chang, L.J. & Sanfey, A.G. Expectations in the Ultimatum Game: Distinct effects of mean and variance of expected offers. *Frontiers in Psychology* 9, 992 (2018).
10. Lee, T., et al. Genetic influences on four measures of executive functions and their covariation with general cognitive ability: The Older Australian Twins Study. *Behavior Genetics* 42, 528-538 (2012).
11. Eftedal, N.H., et al. Justice sensitivity is undergirded by separate heritable motivations to be morally principled and opportunistic. *Scientific Reports* 12, 5402 (2022).
12. Wallace, B., Cesarini, D., Lichtenstein, P. & Johannesson, M. Heritability of ultimatum game responder behavior. *Proceedings of the National Academy of Sciences* 104, 15631-15634 (2007).
13. Wang, Y., Luo, Y.L., Wu, M.S. & Zhou, Y. Heritability of justice sensitivity. *Journal of Individual Differences* (2022).
14. Wang, Y., et al. Born for fairness: Evidence of genetic contribution to a neural basis of fairness intuition. *Social Cognitive and Affective Neuroscience* 14, 539-548 (2019).
15. Scourfield, J., Martin, N., Lewis, G. & McGuffin, P. Heritability of social cognitive skills in children and adolescents. *The British Journal of Psychiatry* 175, 559-564 (1999).
16. Gu, X., et al. Necessary, yet dissociable contributions of the insular and

ventromedial prefrontal cortices to norm adaptation: Computational and lesion evidence in humans. *Journal of Neuroscience* 35, 467-473 (2015).
17. Hetu, S., Luo, Y., D'Ardenne, K., Lohrenz, T. & Montague, P.R. Human substantia nigra and ventral tegmental area involvement in computing social error signals during the ultimatum game. *Social Cognitive and Affective Neuroscience* 12, 1972-1982 (2017).
18. Xiang, T., Lohrenz, T. & Montague, P.R. Computational substrates of norms and their violations during social exchange. *Journal of Neuroscience* 33, 1099-1108 (2013).
19. Sáez, I., Zhu, L., Set, E., Kayser, A. & Hsu, M. Dopamine modulates egalitarian behavior in humans. *Current Biology* 25, 912-919 (2015).
20. Bogacz, R. Dopamine role in learning and action inference. *eLife* 9, e53262 (2020).
21. Frank, M.J., Moustafa, A.A., Haughey, H.M., Curran, T. & Hutchison, K.E. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences* 104, 16311-16316 (2007).
22. Klein, T.A., et al. Genetically determined differences in learning from errors. *Science* 318, 1642-1645 (2007).
23. Pohjalainen, T., et al. The A1 allele of the human D2 dopamine receptor gene predicts low D2 receptor availability in healthy volunteers. *Molecular Psychiatry* 3, 256-260 (1998).
24. Zhang, Y., et al. Polymorphisms in human dopamine D2 receptor gene affect gene expression, splicing, and neuronal activity during working memory. *Proceedings of the National Academy of Sciences* 104, 20552-20557 (2007).
25. Smith, C., et al. The impact of common dopamine D2 receptor gene polymorphisms on D2/3 receptor availability: C957T as a key determinant in putamen and ventral striatum. *Translational Psychiatry* 7, e1091-e1091 (2017).
26. Den Ouden, H.E., et al. Dissociable effects of dopamine and serotonin on reversal learning. *Neuron* 80, 1090-1100 (2013).
27. Rogers, R.D. The roles of dopamine and serotonin in decision making: Evidence from pharmacological experiments in humans. *Neuropsychopharmacology* 36, 114-132 (2011).
28. Crockett, M.J., Clark, L., Hauser, M.D. & Robbins, T.W. Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences* 107, 17433-17438 (2010).
29. Crockett, M.J., Clark, L., Tabibnia, G., Lieberman, M.D. & Robbins, T.W. Serotonin modulates behavioral reactions to unfairness. *Science* 320, 1739-1739 (2008).
30. Emanuele, E., Brondino, N., Bertona, M., Re, S. & Geroldi, D. Relationship between platelet serotonin content and rejections of unfair offers in the ultimatum game. *Neuroscience Letters* 437, 158-161 (2008).
31. Takahashi, H., et al. Honesty mediates the relationship between serotonin

and reaction to unfairness. *Proceedings of the National Academy of Sciences* 109, 4281-4284 (2012).
32. Chen, G.-L., Vallender, E.J. & Miller, G.M. Functional characterization of the human TPH2 5' regulatory region: Untranslated region and polymorphisms modulate gene expression in vitro. *Human Genetics* 122, 645-657 (2008).
33. Scheuch, K., et al. Characterization of a functional promoter polymorphism of the human tryptophan hydroxylase 2 gene in serotonergic raphe neurons. *Biological Psychiatry* 62, 1288-1294 (2007).
34. Steenbergen, L., Jongkees, B.J., Sellaro, R. & Colzato, L.S. Tryptophan supplementation modulates social behavior: A review. *Neuroscience & Biobehavioral Reviews* 64, 346-358 (2016).
35. Gutknecht, L., et al. Tryptophan hydroxylase-2 gene variation influences personality traits and disorders related to emotional dysregulation. *International Journal of Neuropsychopharmacology* 10, 309-320 (2007).
36. Gao, J., et al. TPH2 gene polymorphisms and major depression—A meta-analysis. *PLOS ONE* 7, e36721 (2012).
37. Gärtner, A., Strobel, A., Reif, A., Lesch, K.-P. & Enge, S. Genetic variation in serotonin function impacts on altruistic punishment in the ultimatum game: A longitudinal approach. *Brain and Cognition* 125, 37-44 (2018).
38. Fehr, E. & Camerer, C.F. Social neuroeconomics: The neural circuitry of social preferences. *Trends in Cognitive Sciences* 11, 419-427 (2007).
39. Hawrylycz, M., et al. Canonical genetic signatures of the adult human brain. *Nature Neuroscience* 18, 1832-1844 (2015).
40. Rao, L.-L., Zhou, Y., Zheng, D., Yang, L.-Q. & Li, S. Genetic contribution to variation in risk taking: A functional MRI twin study of the balloon analogue risk task. *Psychological Science* 29, 1679-1691 (2018).
41. Beck, A.T., Steer, R.A. & Brown, G.K. *Beck Depression Inventory–II (BDI-II)* (1996).
42. Montague, P.R. & Lohrenz, T. To detect and correct: Norm violations and their enforcement. *Neuron* 56, 14-18 (2007).
43. Bellucci, G., Feng, C., Camilleri, J., Eickhoff, S.B. & Krueger, F. The role of the anterior insula in social norm compliance and enforcement: Evidence from coordinate-based and functional connectivity meta-analyses. *Neuroscience & Biobehavioral Reviews* 92, 378-389 (2018).
44. Civai, C., Crescentini, C., Rustichini, A. & Rumiati, R.I. Equality versus self-interest in the brain: Differential roles of anterior insula and medial prefrontal cortex. *NeuroImage* 62, 102-112 (2012).
45. Krueger, F., Grafman, J. & McCabe, K. Neural correlates of economic game playing. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363, 3859-3874 (2008).
46. Chang, L.J. & Sanfey, A.G. Great expectations: Neural computations underlying the use of social norms in decision-making. *Social Cognitive and Affective Neuroscience* 8, 277-284 (2013).
47. Shaw, D.J., et al. A dual-fMRI investigation of the iterated Ultimatum

Game reveals that reciprocal behaviour is associated with neural alignment. *Scientific Reports* 8, 10896 (2018).
48. Gabay, A.S., Radua, J., Kempton, M.J. & Mehta, M.A. The Ultimatum Game and the brain: A meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews* 47, 549-558 (2014).
49. Aquino, T.G., Cockburn, J., Mamelak, A.N., Rutishauser, U. & O'Doherty, J.P. Neurons in human pre-supplementary motor area encode key computations for value-based choice. *Nature Human Behaviour*, 1-16 (2023).
50. Wunderlich, K., Rangel, A. & O'Doherty, J.P. Neural computations underlying action-based decision making in the human brain. *Proceedings of the National Academy of Sciences* 106, 17199-17204 (2009).
51. Bonini, F., et al. Action monitoring and medial frontal cortex: Leading role of supplementary motor area. *Science* 343, 888-891 (2014).
52. Schultz, W. Updating dopamine reward signals. *Current Opinion in Neurobiology* 23, 229-238 (2013).
53. Hayden, E.P., et al. The dopamine D2 receptor gene and depressive and anxious symptoms in childhood: Associations and evidence for gene-environment correlation and gene-environment interaction. *Psychiatric Genetics* 20, 304-310 (2010).
54. Ike, K.G.O., et al. The human neuropsychiatric risk gene Drd2 is necessary for social functioning across evolutionary distant species. *Molecular Psychiatry* 29, 518-528 (2024).
55. Ding, Q., et al. Brain network integration underpins differential susceptibility of adolescent anxiety. *Psychological Medicine*, 1-10 (2023).
56. Montalto, A., et al. Negative association between anterior insula activation and resilience during sustained attention: An fMRI twin study. *Psychological Medicine* 53, 3187-3199 (2023).
57. Zheng, D., Chen, J., Wang, X. & Zhou, Y. Genetic contribution to the phenotypic correlation between trait impulsivity and resting-state functional connectivity of the amygdala and its subregions. *NeuroImage* 201, 115997 (2019).
58. Chen, J., et al. The Beijing Twin Study (BeTwiSt): A longitudinal study of child and adolescent development. *Twin Research and Human Genetics* 16, 91-97 (2013).
59. Jin, Y., et al. The perception-behavior dissociation in the ultimatum game in unmedicated patients with major depressive disorders. *Journal of Psychopathology and Clinical Science* 131, 253 (2022).
60. Gagne, C., Zika, O., Dayan, P. & Bishop, S.J. Impaired adaptation of learning to contingency volatility in internalizing psychopathology. *eLife* 9, e61387 (2020).
61. Carpenter, B., et al. Stan: A probabilistic programming language. *Journal of Statistical Software* 76 (2017).
62. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* 27, 1413-1432 (2017).

63. R Core Team. *R: A language and environment for statistical computing* (2013).
64. Akaike, H. Information theory and an extension of the maximum likelihood principle. *Selected Papers of Hirotugu Akaike*, 199-213 (1998).
65. Kline, R.B. *Principles and Practice of Structural Equation Modeling* (Guilford Publications, 2023).
66. Friston, K., et al. Bayesian decoding of brain images. *NeuroImage* 39, 181-205 (2008).
67. Loehlin, J.C. The Cholesky approach: A cautionary note. *Behavior Genetics* 26, 65-69 (1996).
68. Hawrylycz, M.J., et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489, 391-399 (2012).
69. Arnatkevičiūtė, A., Fulcher, B.D. & Fornito, A. A practical guide to linking brain-wide gene expression and neuroimaging data. *NeuroImage* 189, 353-367 (2019).
70. Arloth, J., Bader, D.M., Röh, S. & Altmann, A. Re-annotator: Annotation pipeline for microarray probe sequences. *PLOS ONE* 10, e0139516 (2015).
71. Markello, R.D., et al. Standardizing workflows in imaging transcriptomics with the abagen toolbox. *eLife* 10, e72129 (2021).
72. Shannon, P., et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* 13, 2498-2504 (2003).

Table 1. Intra-class correlations, model comparison, and heritability estimates for model parameters and acceptance rates

Variable	ICC MZ [95% CI]	ICC DZ [95% CI]	Fisher z, p	Model	A [95% CI]	C [95% CI]	E [95% CI]	AIC
Initial fairness norm	.62** [.32, .79]	[-.43, .57]	z = 2.31, p = .010	**AE**	.41 [.18, .59]	.00 [.00, .41]	.59 [.41, .76]	65
Learning rate (human)	.57** [.23, .76]	[-.22, .64]	z = 1.35, p = .088	**AE**	.36 [.13, .55]	.10 [.00, .46]	.54 [.45, .87]	68
Fairness sensitivity (human)	.54** [.17, .74]	[-.47, .56]	z = 1.85, p = .032	**AE**	.32 [.08, .53]	.00 [.00, .37]	.68 [.47, .92]	68
Acceptance rate (human)	.56** [.22, .76]	[-.29, .62]	z = 1.54, p = .062	**AE**	.37 [.14, .56]	.00 [.00, .44]	.63 [.44, .86]	63

Variable	ICC MZ [95% CI]	ICC DZ [95% CI]	Fisher z, p	Model	A [95% CI]	C [95% CI]	E [95% CI]	AIC
Fairness sensitivity (computer)	.53** [.15, .74]	[-.26, .63]	z = 1.20, p = .114	ACE	.25 [.00, .55]	.32 [.00, .53]	.43 [.32, .63]	64
Acceptance rate (computer)	.52** [.14, .73]	[-.24, .63]	z = 1.09, p = .138	ACE	.00 [.00, .41]	.37 [.00, .56]	.63 [.44, .86]	64

Note: Optimal models are bolded and underlined. Significant ICCs ($p < .001$) are denoted by **. ACE: genetic (A), shared environmental (C), and non-shared environmental (E) contributions; AIC: Akaike Information Criterion; DZ: dizygotic pairs; ICC: intraclass correlation; MZ: monozygotic pairs.

Table 2. The relationship between SNPs, SMA/mSFG activity, and the learning rate

Regressor	Correlation with learning rate	Correlation with SMA/mSFG activity
Dopaminergic Genes		
rs1800497	F(1, 137.86) = 3.77†, p = .054Learning rate of GG > TT/GT	F(1, 138.87) = 4.92, p = .028GG had larger magnitude of negative SMA/mSFG activation than TT/GT
rs2283265	F(1, 148.99) = 2.91†, p = .090Learning rate of CC > TT/CT	F(1, 147.85) = 6.85, p = .010CC had larger magnitude of negative SMA/mSFG activation than TT/CT
Dopaminergic additive score (rs1800497 + rs2283265)	F(1, 143.93) = 3.57†, p = .061More C-G pair associated with higher learning rates	F(1, 144.09) = 6.17, p = .014More C-G pair associated with larger magnitude of negative activation in SMA/mSFG
rs6277	F(1, 154.24) = .46, p = .501	F(1, 150.94) = 1.20, p = .275
TPH2 (rs4570625)	F(1, 127.36) = .98, p = .323	F(1, 125.38) = .08, p = .772

Regressor	Correlation with learning rate	Correlation with SMA/mSFG activity
Dopaminergic Genes and TPH2		
Dopaminergic additive score	F(1, 136.01) = 5.69, p = .018	F(1, 135.62) = 5.71, p = .018
TPH2	F(1, 151.73) = .25, p = .616	F(1, 151.31) = 1.02, p = .315
Interaction	F(1, 165.75) = 2.60, p = .109	F(1, 165.44) = .07, p = .793

Note: Significant results are bolded. Marginally significant results are marked with †.

Table 3. The relationship between SNPs and the initial fairness norm and fairness sensitivity

Regressor	Correlation with initial fairness norm	Correlation with fairness sensitivity
Dopaminergic Genes		
rs1800497	F(1, 141.41) = 4.71, p = .032Initial fairness norm of TT/GT > GG	F(1, 133.42) = 6.33, p = .013Fairness sensitivity of TT/GT > GG
rs2283265	F(1, 153.96) = 3.83†, p = .052Initial fairness norm of TT/CT > CC	F(1, 145.49) = 6.06, p = .015Fairness sensitivity of TT/CT > CC
Dopaminergic additive score	F(1, 148.43) = 4.42, p = .037More C-G pair associated with lower fairness norms	F(1, 139.65) = 6.48, p = .012More C-G pair associated with lower fairness sensitivity
rs6277	F(1, 155.95) = .05, p = .816	F(1, 149.07) = .08, p = .775
TPH2 (rs4570625)	F(1, 149.21) = 6.75, p = .010Initial fairness norm of GG/GT > TT	F(1, 143.08) = 2.69, p = .103
Dopaminergic Genes and TPH2		
TPH2	F(1, 153.25) = 7.30, p = .008	F(1, 146.94) = 3.77, p = .054
Dopaminergic additive score	F(1, 137.68) = 5.76, p = .018	F(1, 130.71) = 7.30, p = .008

Regressor	Correlation with initial fairness norm	Correlation with fairness sensitivity
Interaction	F(1, 166.63) = .33, p = .566	F(1, 162.60) = .34, p = .561

Note: Significant results are bolded. Marginally significant results are marked with †.

Fig. 1 Analytical Procedures. (A) Computational modeling of the decision-making process in the UG and heritability analysis with univariate genetic modeling on key parameter indices (i.e., learning rate, initial fairness norm, and fairness sensitivity) and behavioral indices (i.e., acceptance rate). (B) Construction of 1st-level GLM and voxel-wise heritability analysis to identify brain clusters with credible genetic influence. (C) Gene-brain-behavior association analysis: bivariate genetic modeling analysis to identify brain clusters showing both significant phenotypic and genetic correlation with the learning rate, and mediation analysis to examine the mediating role of neural encoding of PE in SMA/mSFG between DRD2 polymorphisms and learning capability, and gene enrichment analysis of SMA/mSFG to provide supplementary evidence for high expression of dopaminergic genes in this brain region. (D) Learning capability could buffer against the occurrence of long-term depressive symptoms, and we investigated whether DRD2 constituted their common genetic basis, and further tested the mediating role of learning capability between DRD2 gene polymorphism and depressive symptoms.

Fig. 2 Experimental Procedures. (A) In each trial, participants were randomly matched to either an anonymous human proposer or a computer-generated proposer. Each human proposer was denoted by a random alphanumeric code instead of names or photos to control confounding effects. In the cover story, participants were told that proposals from human proposers were made by real participants, whereas proposals from computer proposers were randomly generated by a computer program. Proposals from human/computer proposers were identical and preset by experimenters. Participants could accept or reject the proposal by pressing different buttons on an MRI-compatible button box. Decision "acceptance" resulted in splits according to the proposal, whereas "rejection" resulted in nothing for both players. After making the decision, results were shown on the screen. (B) Example experimental procedure for a single shot of the game where participants faced a human proposer. (C) Participants received proposals with fluctuating fairness levels from different proposers in different shots of the game.

Fig. 3 The 18 ROIs extracted from the PPM and their mean heritability across voxels.

Fig. 4 The correlation between brain activity modulated by norm PE and the learning rate. (A) The correlation between SMA/mSFG encoding of PE and the learning rate. (B) The correlation between right anterior insula encoding of

PE and the learning rate.

Fig. 5 The gene-brain-behavior mediation models. (A) The mediation of SMA/mSFG encoding of PE on the relationship between rs1800497 and the learning rate. (B) The mediation of SMA/mSFG encoding of PE on the relationship between rs2283265 and the learning rate. (C) The mediation of SMA/mSFG encoding of PE on the relationship between dopaminergic additive scores and the learning rate.

Fig. 6 Metascape enrichment analysis. (A) Metascape enrichment analysis on top expressed genes in the SMA/mSFG. (B) Metascape enrichment network visualization of top 20 genes.

Note: Figure translations are in progress. See original paper for figures.