Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Behavioral Sciences

A neuroeconomic framework for investigating gender disparities in moralistic punishment

Megha Chawla, Brian D Earp and Molly J Crockett

Moralistic punishment is common in humans and functions to discourage perceived moral transgressions. Research in neuroeconomics suggests that moralistic punishment behavior is associated with activity in neural systems involved in detecting norm violations and in value-based decision-making. Separately, research in philosophy and social psychology highlights different moral expectations for girls/women and boys/men. Here, we synthesize these perspectives to propose a framework for investigating gender disparities in punishment. We propose such disparities may arise through multiple channels, including (1) differences in how the neural salience network responds to perceived norm violations, with stronger responses when women (versus men) violate feminine-coded norms, and when men (versus women) violate masculine-coded norms; and (2) differences in how the neural valuation network tracks the value of punishment decisions, with stronger responses when punishing gender-specific norm violations. We review literature on gendered moral expectations and neural mechanisms underlying moralistic punishment, and suggest hypotheses for future research.

Address
Department of Psychology, Yale University, United States

Corresponding author: Crockett, Molly J (molly.crockett@yale.edu)

Current Opinion in Behavioral Sciences 2020, **34**:166–172

This review comes from a themed issue on **Emotion, motivation, personality and social sciences *Political Ideologies***

Edited by **John Jost**, **Eran Halperin** and **Kristin Laurin**

https://doi.org/10.1016/j.cobeha.2020.03.011

Introduction

Humans often punish others who are perceived to violate moral norms: socially prescribed rules for how one ought to behave.[1] An extensive literature in moral psychology and behavioral economics has characterized the psychological mechanisms of such moralistic punishment, and more recent work in neuroeconomics has begun to describe its underlying neural mechanisms [1]. For the most part, this work has focused on anonymous social interactions, where the punisher knows little to nothing about the target of punishment apart from their evidently having committed a norm violation. However, research in social psychology and philosophy suggests that punishment often depends on targets' social-identity features. For example, racial and gender disparities have been documented in the content or application of moral norms, and, more generally, moral judgments are highly sensitive to relational context [2]. However, existing research has still only scratched the surface of socially situated moralistic punishment, including its underpinnings in the brain. Some recent work has begun to explore the neural basis of racial disparities in punishment [3], but so far, gender disparities have received little attention.

Here, we synthesize work in neuroeconomics, philosophy and social psychology to explore when and how different expectations about moral behavior for girls/women and boys/men — or *gendered moral norms* — affect decisions to punish perceived moral norm violations. In what follows, we sketch out a testable framework for understanding the psychology and neuroscience of gender disparities in moralistic punishment.

Moralistic punishment and its neural basis

Moralistic punishment is the intentional application of an aversive consequence to someone who is perceived to have violated a moral norm. Within psychology, there is a rich tradition of studying social norms more generally as well as their influence on individual judgments and behavior [4–6]. Here, we focus on moral norms in particular, and how people respond to their perceived violation. Moralistic punishment can take many forms, including private or public shaming, physical aggression, or social exclusion. In laboratory studies focused on fairness or reciprocity norms primarily, punishment is typically measured using economic games where punishers can reduce the payoff of norm violators. Though moral norms and associated punishments do vary cross-culturally, punishment behavior of one kind or another is a cultural universal [7], and emerges early in human development [8–10].

Witnessing or experiencing a moral norm transgression, such as a violation of fairness, is affectively aversive. Unfairness in economic games has been found to incite feelings of anger, sadness, disappointment, and spite

[1] We use the terms 'moral' and 'moralistic' throughout the manuscript as descriptive terms, referring to attitudes and judgments of ordinary people that have moral content. In other words, we are not making normative claims about what is in fact morally right or wrong.

toward the unfair player [11•,12,13]. Negative affective responses following norm violations are highly predictive of punishment behavior [12], and may partly explain why punishment behavior linearly increases with the unfairness of the offer proposed [14], an effect that is largely consistent across cultures [7]. Likewise, intentional norm violations are punished more severely than unintentional norm violations [15,16], perhaps in part because the former are more upsetting than the latter.

Research on the neural mechanisms of moralistic punishment has identified two distinct stages of this behavior, each of which engages a different neural circuit (see Ref. [1] for a meta-analysis). The first stage involves *detecting* a norm violation, which broadly engages the salience network. The second stage involves *deciding* to punish, which engages regions implicated in value-based decision-making.

Detecting a moral norm violation, such as an unfair offer in an economic exchange, activates several regions including the anterior insula, dorsal anterior cingulate cortex (dACC), superior temporal sulcus (including temporoparietal junction) and ventrolateral, dorsolateral, and dorsomedial prefrontal cortex (PFC) [1,16,17,18••]. The anterior insula and dACC are core hubs of the salience network and are involved in integrating sensory, affective, and cognitive information in order to direct attention to salient stimuli [19]. Activation of the anterior insula in particular has been associated with signaling inequality [20], and is parametrically correlated with punishment behavior in economic games [17,21,22]. Given the insula's role in emotional reactivity and salience attribution, activation of the anterior insula might signal an affective response to the norm violation.

At the detection stage, the 'norm prediction error' has been proposed as a mechanism guiding punishment decisions [23]. Norm prediction errors arise when there is a mismatch between the internal representation of the norm (modelled as a probability distribution over possible behaviors) and the observed behavior [24]. A Bayesian ideal observer model can be used to determine a norm prediction error parameter, which has been found to be encoded in the insula, ventral striatum and ventromedial PFC (vmPFC) [24]. In a key study, this prediction error signal correlated with participants' subjective feelings about unfair offers and vmPFC activation. Importantly, in an economic exchange experiment, participants who had higher expectations of fairness reported more negative affective responses to fairness violations and were more likely to punish those violations than participants who had lower expectations of fairness. Such findings point to a close relationship between detecting norm violations (at least those relating to fairness), experiencing negative affect, and punishment behavior.

With respect to the decision stage, behavioral studies suggest that punishment is generally satisfying to the punisher [25–27], while neuroimaging studies implicate valuation circuitry in punishment decisions. When participants engage in moralistic punishment, there is increased activation in the dorsal and ventral striatum, dorsolateral PFC, and vmPFC [15,28–30], all of which are implicated in value-based decision-making [31] and reinforcement learning [32]. This neural evidence, in concert with behavioral and self-report data from participants, indicates that punishment is motivating to the punisher and deciding to punish engages neural valuation circuitry.

Gender disparities in moral norms

Gender stereotypes include normative expectations for behavior that differ for persons presumed to be male or female[2] [33,34••,35,36,37••]. Seminal work in social psychology has identified prescriptive and proscriptive norms for men and women in Western societies [33,37••,38–41], and these gendered norms have been largely stable over time [37••,42]. Prentice and Carranza's [33] study of gendered traits is especially nuanced and differentiates between prescriptive, proscriptive, and descriptive norms for women and men. Prescriptions that are intensified for women include being cooperative, warm, kind, friendly, supportive, and nurturing, while intensified proscriptions include traits like promiscuity and aggression [33,43]. Adherence to these expectations could serve to promote the fulfillment of a caregiving function [44], which, in Western societies, has traditionally been more normatively expected of women than of men, especially in certain (e.g. domestic) domains [45,46]. For men, intensified prescriptions include being aggressive, competitive, and assertive, while intensified proscriptions include emotionality, approval-seeking, and weakness [33]. Adherence to these expectations might serve to promote a hierarchical function [44], which, in patriarchal social orders, normatively positions men in a dominant role (by definition).

This asymmetric positioning may help to explain the different moral norms that are often applied to men and women. Indeed, moralizing behavior in the first place (that is, treating otherwise neutral or inoffensive behavior as a matter of moral concern) also often differs between genders [47], such that the behavior may be judged to be normal or natural for one gender, but morally inappropriate for the other. One theory holds that gender systems within patriarchal societies — that is, societies organized along hierarchical lines with men normatively expected to occupy dominant roles and women normatively expected

[2] Because past work on gender stereotypes has focused primarily on binary definitions of sex or gender, the current review cannot effectively assess disparities in punishment that may arise in relation to intersex, transgender, and/or non-binary identities, which is a major limitation. Ideally, future work will be able to consider disparities beyond the sex/gender binary.

to occupy subordinate roles — serve to uphold this male/female hierarchy as the default basis for social coordination (for a classic overview, see Ref. [48]). The philosopher Kate Manne has recently proposed that misogyny is one powerful mechanism by which this hierarchy is enforced. In particular, misogyny targets women's perceived behavioral deviations from the norms that govern their prescribed roles (detection stage) and encouraging their punishment accordingly (decision stage) [34••]. These norms include expectations that women will provide particular moral goods to men such as attention, love, care, compassion, and honesty. Conversely, women are expected not to be agentic, power-hungry, dominant, or to seek the valuable roles that are normatively expected to be fulfilled by men.

While Manne focuses more on the social functions of misogyny than on its psychological nature, she explicitly highlights parallels between misogyny and moralistic punishment:

> "*If [misogyny] feels like anything at all, it will tend to be righteous: like standing up for oneself or for morality, or – often combining the two – for the 'little guy'. It often feels to those in its grip like a moral crusade, not a witch hunt. And it may pursue its targets not in the spirit of hating women but, rather, of loving justice.*"

Manne's claim here is an empirical one, and, as such, requires testing to assess its generalizability. In one perspective, justice sensitivity may be negatively associated with gender-specific system justification [49]. However, there is also evidence that violent or oppressive behavior is often motivated by, and experienced as, morally righteous [50]. Together with the theoretical and empirical work described earlier, then, a framework begins to emerge according to which perceived violations of gendered moral norms should elicit different patterns of moralistic punishment for men and women in both behavior and brain activity. Specifically, violations of feminine-coded prescriptive and proscriptive norms are expected to be more harshly punished when women commit these transgressions (e.g. fail to be caring), and violations of masculine-coded prescriptions and proscriptions are expected to be more harshly punished when committed by men (e.g. fail to be courageous in the face of a perceived threat). When people detect that a norm violation has occurred, neural activity in the salience network should correspond to extent of the disparity between the normative expectation and the observed behavior (see Figure 1a). Accordingly, there should be stronger activity when women (versus men) violate feminine-coded norms, and when men (versus women) violate masculine-coded norms (see Figure 1b). A similar pattern should emerge within the valuation network during decisions to punish, such that network activity should be higher when punishing women than men for violating feminine-coded norms, and higher when punishing men than women for violating masculine-coded norms.

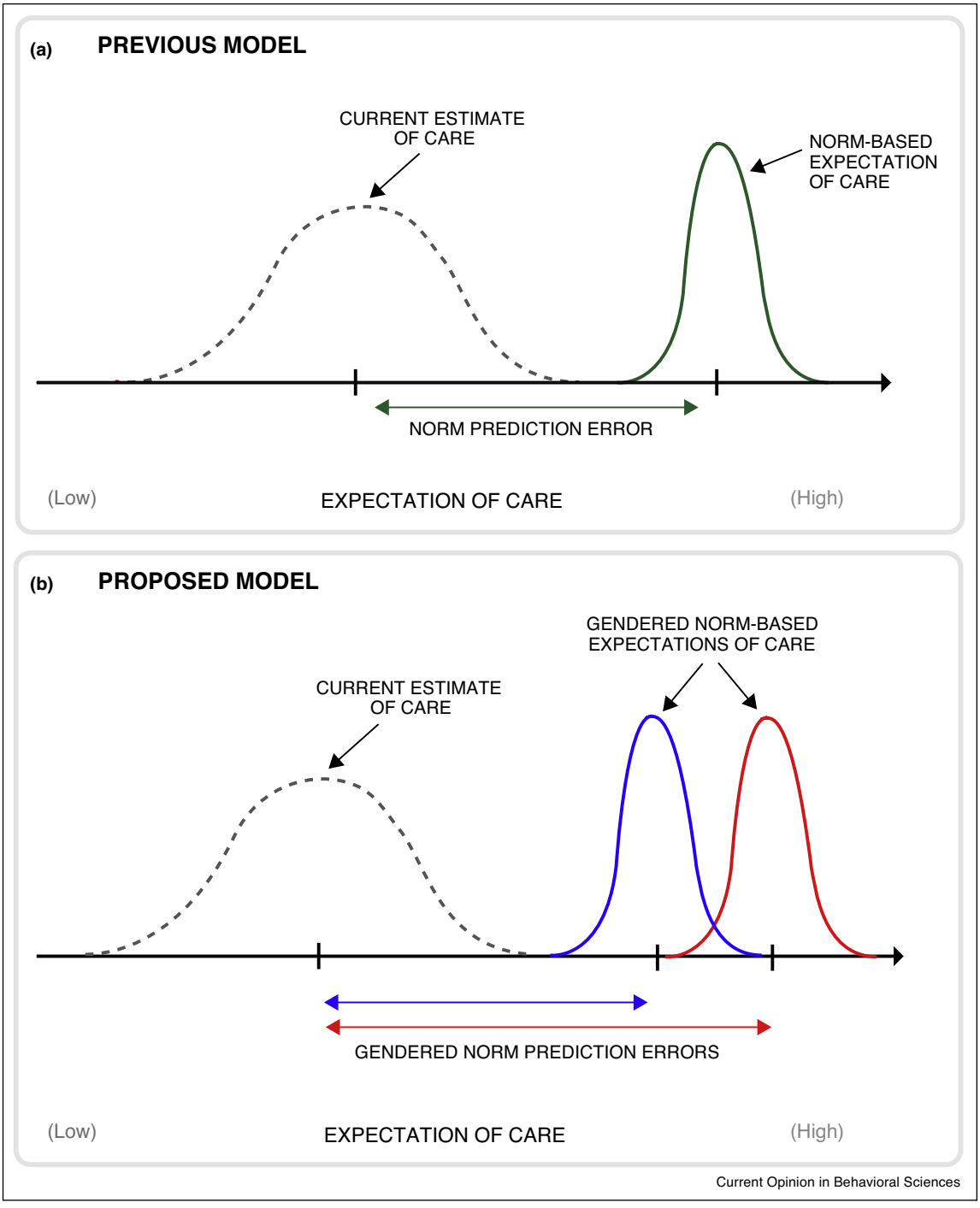
Current evidence for gender disparities in moralistic punishment

Preliminary evidence for our hypothesis comes from the social psychology literature. This literature suggests that both women and men who are deficient in adhering to prescribed gendered moral norms (or who exhibit proscribed gendered moral traits) are judged more harshly by peers than those who are not seen to be deficient in these respects [51,52]. For example, aggressive-independent women who violate prescriptive subordinance norms within hierarchical gendered systems, and passive-dependent men who violate prescriptive dominance norms within such systems, were perceived as less likeable than their norm-conforming counterparts [51]. In other studies, parent dyads with male primary caregivers were perceived as less likeable than dyads with female primary caregivers [52,53]. Further, women deficient in caregiving traits such as empathic concern and sensitivity to others' feelings were judged to be less likeable, effective, supportive, and normative by other women [54].

Additional evidence for gender disparities in punishment following moral norm violations comes from the organizational behavior literature. Narcissistic female leaders, compared to narcissistic male leaders, who lack prescribed traits such as kindness and warmth or possess proscribed traits such as arrogance, were perceived as less effective leaders by subordinates [55]. Kennedy *et al.* leveraged experimental and field data to show that female attorneys were more expected to conform to professional ethics codes and were punished more harshly for ethical transgressions than male attorneys [56]. Echoing this finding, female financial advisers were punished more harshly following misconduct, were more likely to be fired and less likely to find a new job relative to male financial advisers who behaved similarly [57•]. There is also evidence that female leaders were expected to be punished more harshly for their leadership failures than male counterparts in similar positions [58], and ethical failures do more damage to organizational reputations when the organization is led by a woman relative to a man [59].

Behavioral economic research on moralistic punishment has for the most part not considered gender disparities, but there are a few notable exceptions, albeit with mixed results. For example, while one study reported higher rates of punishment of women than men who violate fairness norms in ultimatum games [60], two other studies reported lower levels of punishment for women than for men [61,62]. Finally, a direct replication of one of these studies ([63] replicating [60]) failed to find any gender disparities in punishment.

Figure 1



Gender disparities in the detection of moral norm violations. Here we depict a computational framework for conceptualizing gender disparities in the detection of moral norm violations, using expectations of care as an example context. **(a)** Lohrenz and Montague's model for detecting social norm violations, where norm prediction errors encode discrepancies between (in this example context) observed levels of care and expected levels of care. **(b)** Proposed model for gendered moral norm violations, where gendered moral norms establish different expectations of care on the basis of gender. This results in gendered norm prediction errors according to the (perceived) gender of the actors of observed behavior. Crucially, different expectations of care on the basis of gender result in different prediction errors for the same behavior. To the extent prediction errors drive punishment, there will then be gender disparities in punishment.

One possible explanation for these discrepancies is that the economic games employed in these studies largely invoke moral norms surrounding fairness and reciprocity. However, it is not clear whether violations of reciprocity expectations — as opposed to those concerning care or hierarchy — are especially gendered, even in patriarchal societies. Indeed, reciprocity norms are thought to apply precisely to those situations in which participants are functional equals [44]; and monetary systems may have evolved in part to serve an equalizing function between otherwise unequal groups or individuals with a common interest in fair exchange [64]. As such, the widespread use of economic games to study moralistic punishment behavior is insufficient for fully characterizing its underlying psychology. This is because only reciprocity-based moral norms are at stake in such games, rather than moral norms derived from care, hierarchy, mating, coalition, or other cooperative functions [44], some of which are more likely to be gendered. Future research can inform gender disparities in moral norm violations by examining a broader array of moral norm violations and consequent punishments.

For example, paradigms like the trust game [65] or a modified trust game with antecedent promise stage [66] might probe gender disparities in punishment following violations of trustworthiness and honesty norms. Additionally, vignettes probing moral judgments in a wider array of social and relational contexts might shed further light on gender disparities in blameworthiness following norm violations across a range of interpersonal situations [67].

Conclusions and future directions

Neuroeconomic research on moralistic punishment has identified two distinct stages: detecting that a norm violation has occurred, and deciding to punish the norm violation, which, respectively, engage neural networks involved in salience detection and valuation. Meanwhile, research in social psychology and philosophy show that moral norms vary for men and women, and that men and women are punished for deviating from prescriptive and proscriptive expectations for their respective genders. Synthesizing these distinct literatures, we propose that there are likely to be systematic gender disparities in moralistic punishment, and that violations of gendered moral norms will be associated with neural norm prediction errors at the detection stage, and engagement of the valuation circuitry during the punishment stage in a way that reflects these gender disparities. We stress that existing behavioral economic paradigms may be too narrow in focus to adequately test these predictions, as they are typically concerned with violations of fairness or reciprocity norms only. Since norms surrounding care and hierarchy should, on theoretical grounds, be more likely to be gendered, research paradigms will need to expand to include these norms in order to study gender disparities in moralistic punishment in an ecologically valid manner.

Our proposal raises several new questions that we think warrant investigation. For example, what are the specific gendered moral expectations for men and women in the present day? How do these expectations vary across human societies and demographic groups? Do larger deviations from gendered moral norms result in harsher punishments? How do gender disparities in punishment intersect with disparities on the basis of race, class, sexual orientation, or disability? How does gender non-conformity, along with increasing recognition of gender fluidity and non-binary identities, affect normative expectations? What are the neural mechanisms underlying gender disparities in punishment? Does the brain treat gendered norm violations differently to how it treats other kinds of norm violations? If so, does this happen at the detection stage, the decision stage, or both? Could these neural mechanisms identify measurable targets for interventions to reduce bias? We hope that addressing these questions and others will pave the way toward reducing gendered iniquities in punishment and in society more broadly.

Conflict of interest statement

Nothing declared.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Feng C, Luo Y-J, Krueger F: **Neural signatures of fairness-related normative decision making in the ultimatum game: a coordinate-based meta-analysis**. *Hum Brain Mapp* 2015, **36**:591-602.
2. Clark MS, Lemay EP, Reis HT: **Other people as situations: relational context shapes psychological phenomena**. In *The Oxford Handbook of Situations.* Edited by Rauthmann JF *et al.*: Oxford University Press; 2018.
3. Kubota JT, Banaji MR, Phelps EA: **The neuroscience of race**. *Nat Neurosci* 2012, **15**:940.
4. Sherif M: *The Psychology of Social Norms*. 1936.
5. Schultz PW, Nolan JM, Cialdini RB, Goldstein NJ, Griskevicius V: **The constructive, destructive, and reconstructive power of social norms**. *Psychol Sci* 2007, **18**:429-434.
6. Reno RR, Cialdini RB, Kallgren CA: *The Transsituational Influence of Social Norms*. American Psychological Association; 1993:104-112.
7. Henrich J, McElreath R, Barr A, Ensminger J, Barrett C, Bolyanatz A, Cardenas JC, Gurven M, Gwako E, Henrich N *et al.*: **Costly punishment across human societies**. *Science* , **312**.
8. Salali GD, G.D.SalaliJuda M, M.JudaHenrich J: J. Henrich**Transmission and development of costly punishment in children**. *Evol Hum Behav* 2015, **36**:86-94.
9. Hamlin JK, J.K.HamlinWynn K, K.WynnBloom P, P. BloomMahajan N: N.Mahajan**How infants and toddlers react to antisocial others**. *Proc Natl Acad Sci U S A* 2011, **108**:19931-19936.

10. Yudkin DA, D.A.YudkinVan Bavel JJ, J.J.Van BavelRhodes M: M. Rhodes**Young children police group members at personal cost**. *J Exp Psychol: Gen* 2020, **149**:182-191 http://dx.doi.org/10.1037/xge0000613.

11. Klimecki OM, O.M.KlimeckiSander D, D.SanderVuilleumier P: P. Vuilleumier**Distinct brain areas involved in anger versus punishment during social interactions**. *Sci Rep* 2018, **8**:1-12
This study uses self-report and fMRI measures to demonstrate that self-reported anger following interacting with an unfair agent in an economic game correlates with extent of later punishment of that agent.

12. Pillutla MM, M.M.PillutlaMurnighan JK: J.K.Murnighan**Unfairness, anger, and spite: emotional rejections of ultimatum offers**. *Organ Behav Hum Decis Processes* 1996, **68**:208-224.

13. Steffgen G, G.SteffgenGollwitzer M: M.Gollwitzer*Emotions and Aggressive Behavior*. Hogrefe Publishing; 2007.

14. Fehr E, E.FehrFischbacher U: U.Fischbacher**Third-party punishment and social norms**. *Evol Hum Behav* 2004, **25**:63-87.

15. de Quervain DJ, D.J.de QuervainFischbacher U, U. FischbacherTreyer V, V.TreyerSchellhammer M, M. SchellhammerSchnyder U, U.SchnyderBuck A, A.BuckFehr E: E. Fehr**The neural basis of altruistic punishment**. *Science* 2004, **305**:1254-1258.

16. Zhong S, S.ZhongChark R, R.CharkHsu M, M.HsuChew SH: S.H. Chew**Computational substrates of social norm enforcement by unaffected third parties**. *NeuroImage* 2016, **129**:95-104.

17. Sanfey AG, A.G.SanfeyRilling JK, J.K.RillingAronson JA, J.A. AronsonNystrom LE, L.E.NystromCohen JD: J.D.Cohen**The neural basis of economic decision-making in the ultimatum game**. *Science* 2003, **300**:1755-1758.

18. Cheng X, X.ChengZheng L, L.ZhengLi L, L.LiZheng Y, Y. ZhengGuo X, X.GuoYang G: G.Yang**Anterior insula signals inequalities in a modified ultimatum game**. *Neuroscience* 2017, **348**:126-134
In this study employing fMRI during a modified ultimatum game, the authors find higher BOLD response of the left anterior insula, dACC/dmPFC when receiving unfair and unequal offers from others, highlighting the role of these hubs of the salience network in detecting norm violations. The anterior insula in particular showed high sensitivity to fairness related norms, with higher BOLD activation when inequality between self and others was high, regardless of the direction of the direction of inequality.

19. Menon V: V.Menon**Salience network**. In *Brain Mapping: An Encyclopedic Reference*, , vol 2. Edited by Toga AWA.W.Toga *Brain Mapping: An Encyclopedic Reference*. . Academic Press: Elsevier; 2015:597-611.

20. Hsu M, Anen C, Quartz SR: **The right and the good: distributive justice and neural encoding of equity and efficiency**. *Science* 2008, **320**:1092-1095.

21. Kirk U, Downar J, Montague PR: **Interoception drives increased rational decision-making in meditators playing the ultimatum game**. *Front Neurosci* 2011, **5**.

22. Civai C, Crescentini C, Rustichini A, Rumiati RI: **Equality versus self-interest in the brain: differential roles of anterior insula and medial prefrontal cortex**. *NeuroImage* 2012, **62**:102-112.

23. Montague PR, Lohrenz T: **To detect and correct: norm violations and their enforcement**. *Neuron* 2007, **56**:14-18.

24. Xiang T, Lohrenz T, Montague PR: **Computational substrates of norms and their violations during social exchange**. *J Neurosci* 2013, **33**:1099-1108.

25. Gollwitzer M, Denzler M: **What makes revenge sweet: seeing the offender suffer or delivering a message?** *J Exp Soc Psychol* 2009, **45**:840-844.

26. Gollwitzer M, Meder M, Schmitt M: **What gives victims satisfaction when they seek revenge?** *Eur J Soc Psychol* 2011, **41**:364-374.

27. Crockett MJ, Özdemir Y, Fehr E: **The value of vengeance and the demand for deterrence**. *J Exp Psychol: Gene* 2014, **143**:2279.

28. Strobel A, Zimmermann J, Schmitz A, Reuter M, Lis S, Windmann S, Kirsch P: **Beyond revenge: neural and genetic bases of altruistic punishment**. *NeuroImage* 2011, **54**:671-680.

29. White SF, Brislin SJ, Sinclair S, Blair JR: **Punishing unfairness: rewarding or the organization of a reactively aggressive response?** *Hum Brain Mapp* 2014, **35**:2137-2147.

30. Crockett MJ, Apergis-Schoute A, Herrmann B, Lieberman MD, Müller U, Robbins TW, Clark L: **Serotonin modulates striatal responses to fairness and retaliation in humans**. *J Neurosci* 2013, **33**:3505-3513.

31. Lee D, Seo H: **Neural basis of strategic decision making**. *Trends Neurosci* 2016, **39**:40-48.

32. ODoherty JP, Lee SW, McNamee D: **The structure of reinforcement-learning mechanisms in the human brain**. *Curr Opin Behav Sci* 2015, **1**:94-100.

33. Prentice DA, Carranza E: **What women and men should be, shouldnt be, are allowed to be, and dont have to be: the contents of prescriptive gender stereotypes**. *Psychol Women Q* 2002, **26**:269-281.

34. Manne K: *Down Girl: The Logic of Misogyny*. Oxford University Press; 2017
Manne presents a rigorous theoretical philosophical framework for misogyny, replete with contemporary examples of its permeating consequences in society. She especially highlights the roles of gendered norms for behavior for women that are consequent of patriarchal social systems; and further, emphasizes the roles of reward and punishment in the enforcement, maintenance, and persistence of these norms.

35. Ridgeway CL: *Framed by Gender: How Gender Inequality Persists in the Modern World*. Oxford University Press; 2011.

36. Bem SL: **The measurement of psychological androgyny**. *J Consult Clin Psychol* 1974, **42**:155.

37. Koenig AM: **Comparing prescriptive and descriptive gender stereotypes about children, adults, and the elderly**. *Front Psychol* 2018, **9** 1086–1086
In a series of three studies, Koenig examines prescriptive and descriptive gender stereotypes across developmental stages from toddlerhood to late adulthood, replicating previous findings for gender norms. Specifically, she finds that for females, prescriptions include warmth and communality and proscriptions include include dominance. For males, prescriptions include being agentic and dominant, while proscriptions include emotionality and weakness. These norms hold true for all age groups except for toddlerhood, and are norms are weaker in late adulthood. She also finds that norms for masculinity are more rigid than norms for femininity.

38. Broverman IK, Vogel SR, Broverman DM, Clarkson FE, Rosenkrantz PS: **Sex-role stereotypes: a current appraisal1**. *J Soc Issues* 1972, **28**:59-78.

39. Diekman AB, Eagly AH: **Stereotypes as dynamic constructs: women and men of the past, present, and future**. *Pers Soc Psychol Bull* 2000, **26**:1171-1188.

40. Auster CJ, Ohm SC: **Masculinity and femininity in contemporary American society: a reevaluation using the Bem Sex-Role inventory**. *Sex Roles* 2000, **43**:499-528.

41. Eagly AH: *Sex Differences in Social Behavior: A Social-Role Interpretation*. Psychology Press; 2013.

42. Donnelly K, Twenge JM: **Masculine and feminine traits on the Bem Sex-Role inventory, 1993–2012: a cross-temporal meta-analysis**. *Sex Roles* 2017, **76**:556-565.

43. Eagly AH, Steffen VJ: **Gender and aggressive behavior: a meta-analytic review of the social psychological literature**. *Psychol Bull* 1986, **100**:309.

44. Bugental DB: **Acquisition of the algorithms of social life: a domain-based approach**. *Psychol Bull* 2000, **126**:187.

45. Eagly AH, Crowley M: **Gender and helping behavior: a meta-analytic review of the social psychological literature**. *Psychol Bull* 1986, **100**:283.

46. Monin JK, Clark MS, Lemay EP: **Communal responsiveness in relationships with female versus male family members**. *Sex Roles* 2008, **59**:176.

47. Stokoe EH: **Mothers, single women and sluts: gender, morality and membership categorization in neighbour disputes**. *Feminism Psychol* 2003, **13**:317-344.

48. Moane G: **Hierarchical systems: patriarchy and colonialism**. *Gender and Colonialism*. Springer; 1999:24-54.

49. Jost JT: **A quarter century of system justification theory: questions, answers, criticisms, and societal applications**. *Br J Soc Psychol* 2019, **58**:263-314.

50. Fiske AP, Rai TS: *Virtuous Violence: Hurting and Killing to Create, Sustain, End, and Honor Social Relationships*. Cambridge University Press; 2014.

51. Costrich N, Feinstein J, Kidder L, Marecek J, Pascale L: **When stereotypes hurt: three studies of penalties for sex-role reversals**. *J Exp Soc Psychol* 1975, **11**:520-530.

52. Brescoll VL, Uhlmann EL: **Attitudes toward traditional and nontraditional parents**. *Psychol Women Q* 2005, **29**:436-445.

53. Gaunt R: **Breadwinning moms, caregiving dads: double standard in social judgments of gender norm violators**. *J Fam Issues* 2012, **34**:3-24.

54. Holmstrom AJ, Burleson BR, Jones SM: **Some consequences for helpers who deliver "Cold Comfort": why its worse for women than men to be inept when providing emotional support**. *Sex Roles* 2005, **53**:153-172.

55. De Hoogh AHB, Den Hartog DN, Nevicka B: **Gender differences in the perceived effectiveness of narcissistic leaders**. *Appl Psychol* 2015, **64**:473-498.

56. Kennedy J, McDonnell M-H, Stephens N: *Does gender raise the ethical bar? Exploring the punishment of ethical violations at work. Exploring the Punishment of Ethical Violations at Work (April 25, 2016). Vanderbilt Owen Graduate School of Management Research Paper*. 2016.

57. Egan ML, Matvos G, Seru A: *When Harry Fired Sally: The Double Standard in Punishing Misconduct*. National Bureau of Economic Research; 2017

In this working paper, the authors present evidence supporting a –gender punishment gap in the financial advisory industry, where women are more likely to be punished during hiring and firing decisions following misconduct, even when their misconduct is less costly compared to that of men.

58. Fisk SR, Overton J: **Who wants to lead? Anticipated gender discrimination reduces womens leadership ambitions**. *Soc Psychol Q* 2019, **82**:319-332.

59. Montgomery NV, Cowen AP: **How leader gender influences external audience response to organizational failures**. *J Pers Soc Psychol* 2019.

60. Solnick SJ: **Gender differences in the ultimatum game**. *Econ Inquiry* 2001, **39**:189-200.

61. Solnick SJ, Schweitzer ME: **The influence of physical attractiveness and gender on ultimatum game decisions**. *Organ Behav Hum Decis Process* 1999, **79**:199-215.

62. Eckel CC, Grossman PJ: **Chivalry and solidarity in ultimatum games**. *Econ Inq* 2001, **39**:171-188.

63. McGee P, Constantinides S: **Repeated play and gender in the ultimatum game**. *J Socio-Econ* 2013, **42**:121-126.

64. Lea SE, Webley P: **Money as tool, money as drug: the biological psychology of a strong incentive**. *Behav Brain Sci* 2006, **29**:161-209.

65. Berg J, Dickhaut J, McCabe K: **Trust, reciprocity, and social history**. *Games Econ Behav* 1995, **10**:122-142.

66. Baumgartner T, Fischbacher U, Feierabend A, Lutz K, Fehr E: **The neural circuitry of a broken promise**. *Neuron* 2009, **64**:756-770.

67. Clark MS, Earp BD, Crockett MJ: **Who are "we" and how and why are we cooperating? Insights from social psychology**. *Behav Brain Sci* 2020 http://dx.doi.org/10.1017/S0140525X19002528. in press.