What are Men and Mothers for?
The Causes and Consequences of Functional Reasoning about Social Categories

Emily Foster-Hanson (emily.fosterhanson@princeton.edu)
Department of Psychology, 222 Peretsman Scully Hall
Princeton, NJ 08540 USA

Tania Lombrozo (lombrozo@princeton.edu)
Department of Psychology, 219 Peretsman Scully Hall
Princeton, NJ 08540 USA

Abstract

Do people attribute functions to gendered social categories? (For instance, is there something men or mothers are *for*?) And if so, do such attributions of function have consequences for normative judgments about what members of these social categories *ought* to do? In the current study, participants ($N = 366$) rated their agreement with 15 statements about the "true functions" of different social categories, in triads of matched masculine, feminine, and superordinate categories (e.g., fathers, mothers, and parents). Participants endorsed functional claims more for some social categories (e.g., parents) than others (e.g., kids), and their background beliefs about gender predicted variation in functional reasoning. However, across categories, participants judged that fulfilling true functions was 'natural' for members of the category, and they judged that category members *ought* to fulfill their true functions.

Keywords: social cognition, functional explanations, normative judgments, gender, essentialism

Introduction

From childhood, people tend to think of human-made artifacts as serving some function or purpose (Rips, 1989). Although even young children understand that people make artifacts but not natural kinds (Gelman & Markman, 1987; Springer & Keil, 1989), people sometimes extend functional reasoning to aspects of the natural world (Kelemen, 1999a; Lombrozo & Rehder, 2012): young children avow that lions are for going in the zoo (Kelemen, 1999c), and many adults agree that trees exist to produce oxygen (Kelemen & Rosset, 2009; Kelemen et al., 2013; Roberts et al., 2021). These beliefs about functions seem to arise from people's background beliefs about the origins of different kinds of things and their features. For example, function attributions about artifacts are grounded in background knowledge that humans create artifacts for specific purposes (e.g., people make knives to cut things; Bloom; 1996). Function attributions about the natural world (e.g., that bees exist to make honey) can rest on beliefs in a divine creator, but they can also stem from common attributions of agency to nature and assumptions about natural selection (Barnes et al., 2017; Gregory, 2009; Kelemen, 2004; Kelemen & Rosset, 2009; Kelemen et al., 2013; Lombrozo & Carey, 2006; Lombrozo & Shtulman, 2006; Mayr, 1982; Shtulman, 2006, 2017; Ware & Gelman, 2014).

Regardless of their source, such attributions of function bring with them a normative standard against which category members can be judged: as better or worse at fulfilling their function. This can in turn license beliefs about how things *ought* to be (Aristotle, ca. 350 B.C.E./1996; Foster-Hanson & Lombrozo, 2021; Lane, 2020). For example, believing that knives are for cutting, or that bees are for making honey, can lead people to think there is something wrong with a knife or a bee that fails to fulfill its function (DiYanni & Kelemen, 2005) or even that it is no longer a knife or a bee at all (Rose & Nichols, 2019; 2020). Here we ask whether this form of functional reasoning extends to social categories, and in particular to gender categories such as *mothers* and *men*. Do people tend to think of such categories of humans in terms of functions, such as protecting others, living a fulfilling life, or raising children? And if so, what are the consequences of such functional reasoning for judgments of how members of these categories *ought* to behave? (*Ought* men to protect, and mothers to nurture?)

While a great deal of prior work shows that people explain the intentional behaviors of individuals in terms of functions and goals (e.g., a person might cross the street *to get to the other side*; Heider, 1958; Malle, 2011), less is known about functional reasoning concerning groups of individuals. At the level of the whole species, there is evidence for some functional reasoning (Lewry et al., 2021), in that many people endorse the view that humans as a species exist to serve a function (e.g., to reproduce). Moreover, this belief is associated with judging that it is morally wrong for an individual human to choose not to fulfill this function (e.g., to choose not to reproduce). There is also evidence that at the more fine-grained level of *social roles* (e.g., doctor, comedian, babysitter, and so on), many categories are associated with clear functions in society (e.g., the function of doctors is to help people who are sick). These features play a role in both classification and normative judgments about what people ought to do (e.g., Kalish, 2012; Kalish & Lawson; 2008; Knobe et al., 2013; Del Pinal & Reuter, 2017).

Generalizing from social roles to social categories more broadly, we might expect people to attribute functions to categories such as *men* and *mothers*, and to judge that men and mothers ought to behave in ways that fulfill their corresponding functions. One aim of this paper is to evaluate whether this is indeed the case. A deeper aim, however, is to

better understand the causes and consequences of such attributions of function. For these questions, gender categories are particularly promising for several reasons.

First, there is likely to be variation across gender categories in the extent to which people adopt a functional perspective, and additionally in the functions they ascribe. *Mothers* and *fathers* are relational categories (Gentner & Kurtz, 2005) with plausible functions associated with raising children, but what are the functions of entity categories such as *boys* or *girls*? Variation across categories in functional reasoning could reflect differences in the extent to which people view them as relational. Second, different gender categories also likely vary in how people think of them—and their functions—as biological or "natural." From at least age 3, children and adults tend to believe that gender marks essentially different kinds of people that exist in nature (Gelman & Taylor, 2000), and they expect gender differences in behavior to emerge naturally (Meyer & Gelman, 2016). One hypothesis we explore is that beliefs about the "natural" go hand in hand with beliefs about the "functional," such that (for example) believing it is natural for mothers to care for children predicts viewing caring for children as a function of mothers.

Gender categories are also a useful test case because there is likely to be variation in judgments across the individuals making those judgments. Many forms of sexism are characterized by beliefs that men and women serve distinct functions in society (Glick & Fiske, 1996; 1999), and individual variation in these beliefs is reflected in people's behavior in both their personal (e.g., Eastwick et al., 2006) and professional lives (e.g., Cassidy & Krendl, 2019). Similarly, individual people might vary in how much they think of categories such as *men* and *women* as relational, with complementary roles and functions in society. There is also documented variation in the extent to which different people "essentialize" gender categories (Fast & Olson, 2018; Skewes et al., 2018), and people who hold stronger gender essentialist beliefs exhibit more support for gender discrimination and more backlash towards gender nonconformity (Kray et al., 2017; Skewes et al., 2018). One hypothesis explore is that variation in sexism and gender essentialism predicts variation in functional reasoning about social categories involving gender. Specifically, we hypothesized that sexist and gender essentialist beliefs might be correlated with endorsing *different* functions for men and women (i.e., more gender stratification).

Finally, gender categories are an ideal test case for investigating normative beliefs about how category members ought to behave. There is already evidence to suggest that essentialist beliefs shape normative judgments, for example by absolving actors of blameworthiness for bad actions that are "natural" and therefore deemed beyond their control (including male infidelity or even rape; Dar-Nimrod et al., 2011; Ismail et al., 2012; see also Brescoll & LaFrance, 2004). Is it also the case that functional reasoning drives normative expectations about gender? For instance, are mothers and fathers judged similarly in whether they ought to fulfill their category-specific functions (e.g., raising children or providing for their families)?

In the study reported below, we consider the causes and consequences of functional reasoning about social categories. We focus on categories with masculine and feminine subordinates (such as parents: mothers & fathers; kids: girls & boys) as an ideal test case for socially-consequential categories about which we are likely to see relevant variation in functional reasoning. Using such categories, we ask: (1) Do people think of social categories as having "true functions" and, if so, how do these function attributions vary across different categories? (2) Do attributions of function vary across individuals with different background beliefs and ideologies? (3) Are beliefs about functions correlated with normative judgments about how members of social categories *ought* to be, above and beyond estimations of feature prevalence?

Method

Pretest

In order to generate a list of possible functions associated with different social categories, we first asked a separate group of participants to answer the open-ended question, "What is the true function of [social category]?" We recruited 120 adults from Prolific (54 female, 65 male, 1 non-binary; M_{age} = 36). Participants were tested using Qualtrics and received $0.50 for participating. All study procedures for this and the main study were approved by the Institutional Review Board of the authors' university.

Each participant in the pretest was randomly assigned to generate a list of 1-3 "true functions" for a single social category, out of 12 possible categories (mothers, fathers, parents, women, men, people, females, males, humans, girls, boys, and kids). We selected these categories because they could be structured into four triads, each consisting of one superordinate category (e.g., parents) and matched masculine and feminine subordinate categories (e.g., fathers, mothers). Participants' generated functions were then coded into common themes using keyword search and checked for accuracy by the first author (all data and analysis code for both the pretest and the main study are available on OSF, https://osf.io/nvjxa/). This yielded a list of the most common functions attributed to each social category, which resulted in the set of 15 unique functions used in the main study[1] (given repeated functions across the categories; see Figure 1).

Participants

We calculated the desired sample size for the study by conducting a power analysis for general linear regression using the pwr.f2.test function in the ("pwr" package; Champely et al., 2020), including a moderate effect size and

---

[1] A total of 16 themes emerged in participants' generated functions, however we omitted functions related to God's design as this captured a different type of functional reasoning than the focus here. The remaining 15 functions were included in the main study.

80% power, with the number of covariates as four in order to allow sufficient power to test the hypotheses described in our preregistration (see https://osf.io/3fk7p). This calculation suggested a sample size of 84 participants per triad, so we recruited 93 participants per triad, none of whom participated in the pretest, to allow for possible drops. Of our total recruited sample of 372 participants (across four triads), 2 did not complete the study and 4 were excluded for incorrectly answering more than 1 out of 3 attention check questions, as described in our preregistration plan. This left 366 participants (198 female, 150 male, 11 non-binary; 4 transgender; 1 gender non-conforming; 2 other; M_{age} = 36). Participants were recruited using Prolific and tested using Qualtrics; they were paid $2.50.

Procedure

Participants rated their agreement with a set of statements about the true functions of matched triads of social categories, based on the functions generated by participants in the pretest. For example, participants rated how much they agreed with the statement, "The true function of fathers is to care for children." These statements were rated on a 1-7 Likert scale (1 = "Strongly disagree", 7 = "Strongly agree"). Participants were randomly assigned to make judgments about all three categories within one of the four triads (parents, kids, humans, and people). All participants made judgments about the two subordinate (gendered) categories first, in random order, followed by judgments about the superordinate category.

Participants also rated on 1-7 Likert scales (1) how prevalent they thought fulfilling each function is among members of the category (e.g., "In general, how many fathers do you think care for children?") (2) how natural it is for the category (e.g., "How natural is it for fathers to care for children?"), and (3) how much members of the category ought to fulfill each function (e.g., "How much do you think fathers ought to fulfill each function (e.g., "How much do you think fathers ought to care for children?").

At the end of the study, we also measured participants' ideological beliefs about gender using the Ambivalent Sexism Inventory (Glick & Fiske, 1996; 1999) and Gender Essentialism Scale (Skewes et al., 2018).

Results

We report the subset of pre-registered analyses that bear on our three motivating questions: (1) Whether people think of social categories in functional terms and, if so, how function attributions vary across different categories; (2) How attributions of function vary across individuals; and (3) Whether beliefs about functions predict normative judgments about how members of social categories *ought* to be. Analyses that were not pre-registered are flagged as exploratory.

Function Attributions Vary Across Categories

Figure 1 reports the mean level of endorsement for each function statement, subdivided by triad and by level within

each triad. As the figure shows, function endorsement was variable but high: all categories had items for which mean function endorsement was significantly above the scale mid-point (i.e., "neither agree nor disagree").

To compare overall function endorsement across categories, we analyzed agreement with function statements as the dependent variable in exploratory linear mixed models (with the lme4 package; Bates et al., 2015), including triad and level as predictors and random intercepts for each participant and function. (The order in which participants made judgments about subordinate masculine and feminine categories was not significant in any model and is excluded.) We report the results of Likelihood Ratio Tests. Participants' endorsement of functions varied by both triad and level (i.e., masculine, feminine, superordinate; 2-way interaction, X^2 (6) = 324.17, $p < .001$). The highest overall function ratings were for the category *parents*, and the lowest for *kids* (see Figure 1). Effects of level were inconsistent across triads.
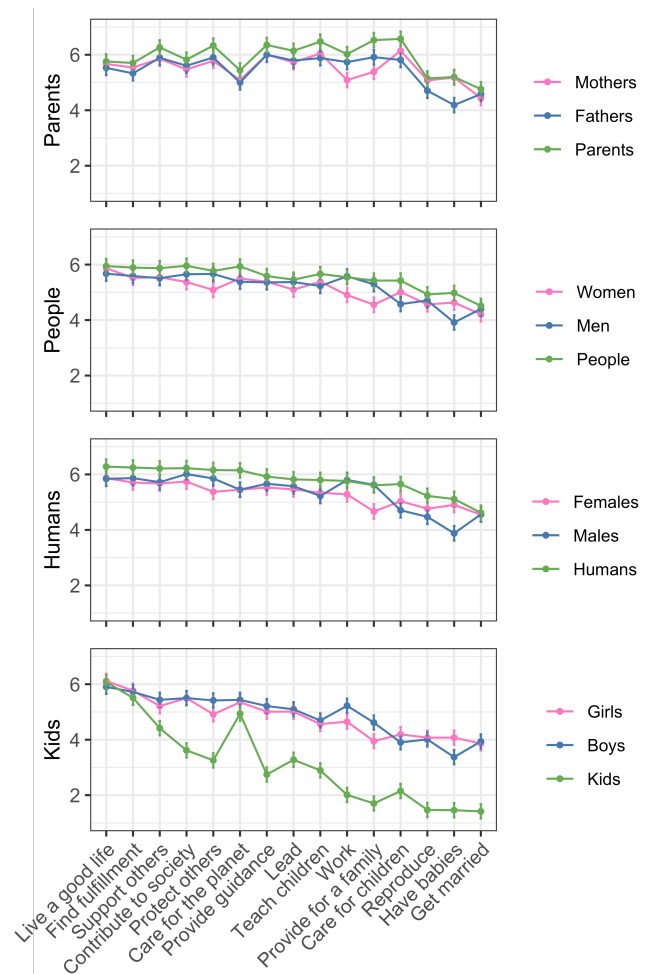

Figure 1: Mean endorsement of "true functions" by category, with 95% CIs. Panels show the four triads.

Functions Vary in Homogeneity Across Categories

In addition to variation in the *strength* of function endorsement, categories might vary in *homogeneity* – that is,

in the extent to which function attributions are consistent across participants. One of our pre-registered hypotheses was that participants' judgments about the functions of mothers would be more homogenous than judgments about any other category. To test this, we compared variances in participants' responses across categories using Levene's tests (with the car package; Fox & Weisberg, 2019). This analysis revealed a significant effect of category on homogeneity, $F(11, 16458) = 102.87$, $p < .001$, with endorsement of the true function of mothers more homogenous ($M_{residual} = 0.89$) than that of all other categories except parents ($M_{residual} = 0.86$; all other categories: $M_{residuals} = 0.92$ to 1.19). Also as predicted, participants' total endorsement of reproductive functions as the "true function" of mothers (top-rated functions: caring for children, teaching children, and providing guidance) was higher than their total endorsement for all other non-reproductive functions ($M_{diff} = 0.15$), but they showed the opposite pattern for every other social category tested—including fathers and parents ($M_{diffs} = -0.07$ to -1.55).

Functions Are Natural

To better understand the sources of function attributions across categories, we performed additional exploratory linear mixed models predicting endorsement of each function statement from ratings of both its prevalence and its naturalness. These models included triad and level as predictors, and random intercepts for both participants and function items. In these analyses, participants' ratings of how "natural" it is to fulfill a function predicted their agreement with the claim that it was the category's "true function" (X^2 (1) = 3067.85, $p < .001$), above and beyond beliefs about the behavior's prevalence (X^2 (1) = 316.32, $p < .001$), across all social categories tested (all simple slopes, $p < .001$). Thus, judgments that fulfilling a given function is "natural" went hand in hand with attributing that function to the category.

Naturalness and Prevalence Vary Across Categories

Across categories, there was also variation in participants' beliefs about the prevalence (2-way triad x level interaction, X^2 (6) = 285.63, $p < .001$), and naturalness of functions (2-way triad x level interaction, X^2 (6) = 444.92, $p < .001$). In follow-up exploratory linear mixed models including only the top three highest-rated functions for each category, and excluding superordinate categories for comparison across gendered categories, participants rated *mothers'* top functions as more prevalent than *fathers'* (pairwise contrast, $p < .001$) but gave similar prevalence estimates for gendered pairs in all other triads (all $ps > .10$; 2-way triad x level interaction, X^2 (3) = 12.15, $p < .001$). Naturalness judgments showed only a main effect of level (X^2 (1) = 6.99, $p = .008$), with participants rating category functions as more natural for feminine categories than masculine categories (pairwise contrast, $p = .008$), and a main effect of triad (X^2 (3) = 4.66, $p = .003$), with participants rating the top functions of *parents* as more natural than the functions of both *kids* (pairwise contrast, $p = .04$) and *people* (pairwise contrast, $p = .007$).

Ideology Predicts Varied Function Attributions

We hypothesized that gender stratification, or overall differences in the endorsement of each function as the true function of paired masculine and feminine categories, would be positively correlated with participants' ideological beliefs including gender essentialism and benevolent sexism. To test this prediction, we computed a gender stratification score for each participant by subtracting their function ratings for the masculine category from their ratings for the feminine category for each function item, then taking the absolute value of the difference to create item-level difference scores per participant. We also averaged participants' responses on the Gender Essentialism Scale (Skewes et al., 2018) and responses to benevolent sexism items from the Ambivalent Sexism Inventory (Fiske & Glick, 1999) into composites. We then analyzed participants' gender stratification scores using backward stepwise regression with the step function (in the lmerTest package; Kuznetsova et al., 2017); the full model included triad and each ideological belief as a main and interactive (with triad) effect and random intercepts for function items.

In the reduced model, there were significant 2-way interactions between triad and gender essentialism, $F(3, 5464) = 5.47$, $p < .001$, and between triad and benevolent sexism, $F(3, 5464) = 7.54$, $p < .001$. Although both gender essentialism and benevolent sexism were positively correlated with gender stratification overall, the different ideological beliefs played different roles across triads (Figure 2). Gender essentialism predicted endorsement of more stratified functions only for boys vs. girls (kids triad, simple slope = .11, $p < .001$) and males vs. females (humans triad, simple slope = .15, $p < .001$; all other $ps < .05$). In contrast, benevolent sexism predicted endorsement of more stratified functions only for men vs. women (people triad, simple slope = .10, $p < .001$; all other $ps < .05$). Participant beliefs did not predict more stratified function attributions for parents, likely due to ceiling effects (see Figure 1). The pattern of results was unchanged if the behavior item "having babies" (which was rarely endorsed for masculine categories) was omitted.
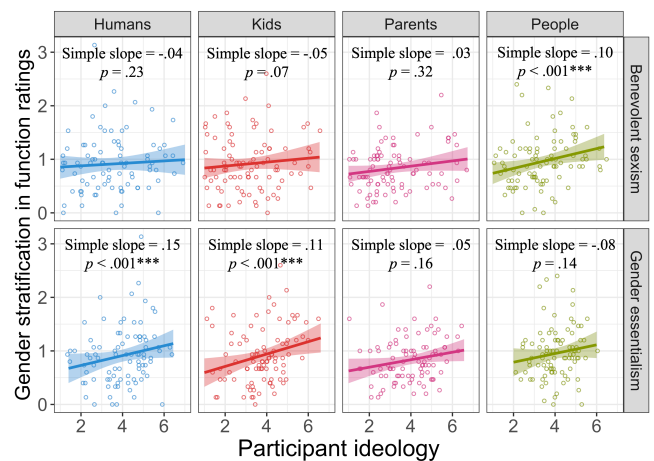

Figure 2: Gender stratification in endorsement of the "true function" of each gendered category, by benevolent sexism

(top row) and gender essentialism (bottom row). Lines are group means with 95% CIs; small circles show individual participant scores. Columns show the four triads.

These results suggest that expecting functions to vary between gender categories because of different biological essences might be distinct from expectations that gender roles entail specific functions in society. We address this interpretation further in the Discussion.

The Consequences of Functional Reasoning for Normative Judgments

Finally, we tested whether endorsement of functional claims predicted normative judgments of behavior. We analyzed participants' agreement with normative statements that category members *ought* to fulfill each function using linear mixed models, testing for main and interactive effects of function endorsement as well as level and triad. We included participants' prevalence estimates as a control, as well as their ideological beliefs (including gender essentialism and benevolent sexism) and random intercepts for participants and function items.
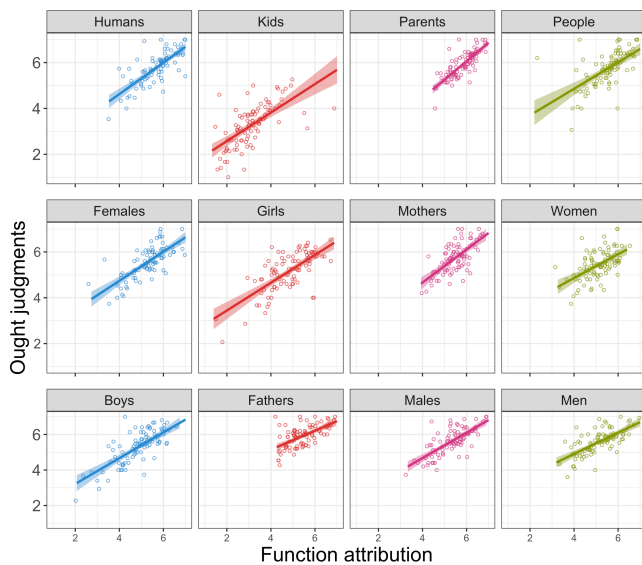
Figure 3: Agreement with *ought* judgements by function attribution for each category. Columns show the four triads, and rows show the three levels (superordinate, feminine, and masculine). Lines are group means with shaded bands showing 95% Confidence Intervals, and small circles are individual participants' average responses. All slopes are significant ($p < .001$).

Overall, participants' beliefs about function predicted their ought judgments (main effect of function ratings, X^2 (1) = 8395.59, $p < .001$) above and beyond estimates of prevalence (main effect of prevalence, X^2 (1) = 916.10, $p < .001$). However, the strength of the relation between beliefs about function and ought judgments varied across categories (3-way function endorsement x level x triad interaction, X^2 (6)

= 5.54, $p < .001$), such that function endorsement was more strongly correlated with ought judgments for some categories (e.g., parents, simple slope = .54, $p < .001$) than for others (e.g., men, simple slope = .41, $p < .001$). However, function ratings were significantly correlated with ought judgments for all categories (all ps < .001; Figure 3). This analysis also revealed a significant main effect of gender essentialism on ought judgments (X^2 (1) = 4.02, $p = .046$), with participants who scored higher on gender essentialism agreeing more strongly overall with normative *ought* judgements.

Discussion

In this study, we examined the causes and consequences of functional reasoning across a range of social categories. We found variation in the strength of function attributions across categories as well as in the homogeneity of those attributions. Despite this variability across categories, participants' ratings for certain functions were above the mid-point for all categories, suggesting that people do think of a wide range of social categories in functional terms. Also across all categories, participants were more likely to attribute functions to categories when they judged them as more "natural" for the category's members, supporting the notion that beliefs about what is natural drive—or at least go hand in hand with—beliefs about the functional.

Participants also varied in the extent to which they viewed different categories in functional terms, with prescriptive implications. For example, participants judged that parents ought to display their functions (caring for children and providing for their families) more than men ought to display their functions (e.g., living a good life; Figure 3). One possible explanation for this difference is that certain behaviors have more obvious consequences for other people within a society. For example, if parents fail to care for their children, then the child will be directly harmed unless other members of society fill the role, but the potential harm of a man failing to live a good life is much more abstract and individual. This variation might also reflect differences in functional roles across relational and entity categories (Kurtz & Gentner, 2005). Participants also endorsed true functions less overall for the superordinate category *kids* than for any other category (Figure 2); it is thus striking that endorsement was much higher for both *boys* and *girls* than for *kids*, suggesting perhaps that gender categories constitute social roles regardless of age.

In addition to these variations across different categories, the current study also found variation in function attributions across individuals with different ideological beliefs. In support of our preregistered hypothesis, gender stratification, or differences in endorsement of the functions of paired masculine and feminine categories (e.g., *women-men*; *mothers-fathers*), varied depending on participants' gender essentialism and benevolent sexism. However, contrary to our prediction, these beliefs played different roles across categories: Participants who scored higher on gender essentialism endorsed more stratified functions for *boys* vs. *girls* and for *males* vs. *females*—categories that people are

more likely to view as natural kinds with features that are determined by biology. In contrast, participants who scored higher on benevolent sexism endorsed more stratified functions for *men* vs. *women*, suggesting that people's function attributions for these categories may stem more from beliefs about agents enacting their social roles than from beliefs about natural tendencies. One interpretation is that benevolent sexism may lead people to view gender categories as relational, while essentialism entails viewing gender categories as distinct entities (Gentner & Kurtz, 2005).

Neither essentialist beliefs nor benevolent sexism predicted function attributions for *parents*, most likely because people endorsed category-specific functions of *mothers*, *fathers*, and *parents* at near-ceiling levels regardless of ideological beliefs. Indeed, endorsement of the functions of *mothers* and *parents* were the most homogenous of all categories tested, in line with our preregistered hypothesis (although, contrary to our prediction, *parents* was more homogenous than *mothers*). Participants also judged that *parents* ought to fulfill their functions more than any other triad, and that fulfilling common functions was more prevalent for *mothers* than *fathers*. Given these mixed results, future research should directly test the causal role of beliefs about naturalness and prevalence for normative judgments of parents. For example, differences in prevalence estimates between mothers and fathers suggest different expectations of conformity; do these expectations lead to different kinds of judgments about mothers and fathers who *fail* to fulfill their functions? Relatedly, the concept *mother* may have a stronger dual character than the concept *father*, with mothers who fail to fulfill their functions more likely to be judged as not "true" mothers (Knobe et al., 2013) or not their "true" selves (Strohminger et al., 2017).

Finally, participants' gender essentialist beliefs predicted stronger agreement with normative ought claims overall. Along with evidence of the central role of naturalness judgments for function ratings, these results hint that beliefs about functions may be a missing link through which essentialism can foster normative expectations of conformity (Haslanger, 2014). Specifically, do people's intuitive explanations for differences in behavior across gender categories entail assuming that they are the product of evolution, like the explanations espoused by some researchers? (e.g., Archer, 1996; Buss, 1995; Tooby & Cosmides, 1992). Given that common folk-biological beliefs about evolution entail thinking that natural selection means improvement over time (Barnes et al., 2017; Gregory, 2009; Kelemen & Rosset, 2009; Kelemen et al., 2013; Lombrozo & Shtulman, 2006; Mayr, 1982; Shtulman, 2006, 2017; Ware & Gelman, 2014), essentialist beliefs about gender could license normative expectations of conformity through the same mechanisms as other is-ought beliefs about natural kinds (Foster-Hanson & Lombrozo, 2021). That is, people might reason that the causal processes that led to current gender differences are inherently beneficent—either because they serve the goals of an intelligent creator, or because of misconceptions that evolution means improvement—so what is common and natural must therefore also be right and good.

This view of gender is compatible with social role theory (e.g., Eagly, 1987; Eagly et al., 2000) because social agents themselves might explain and internalize social stereotypes by incorporating them into their existing causal-explanatory frameworks about the natural world. Along these lines, some researchers have suggested that essentialism itself is a motivated reasoning process (Diesendruck, 2021). These beliefs about what is natural for different gender categories shape both how people judge each other and how they view themselves—including which skills people develop beginning in early childhood (Bian et al., 2017), further perpetuating patterns of gender stratification (Bian et al., 2018; Chestnut et al., 2018; Meyer et al., 2015). Although this way of perpetuating gender stereotypes may be subtle, its effects could be pernicious precisely because its normative entailments are never directly stated, only implied (Becker & Wright, 2011; Haslanger, 2014). Future work should directly test when and how general beliefs about nature mediate the role of essentialism in licensing normative judgments.

The current study was correlational, so future work will be needed to test the causal relation between beliefs about naturalness, attributions of function, and normative judgments. For example, although above we conceptualized naturalness judgments as giving rise to beliefs about function, the reverse relation is also possible: People might expect members of society to fulfill specific functions as the result of collaborative social interaction, and then come to view those functions as natural and inevitable through processes of system justification (Jost & Banaji, 1994; Jost et al., 2004; Kay et al., 2009). The causal direction (naturalness to function or vice versa) could even vary across categories and individuals. Further empirical evidence will be needed to tease apart these possibilities.

It is also unclear how broadly the current results generalize beyond our sample of participants. Although we found substantial variation in ideological beliefs across participants (see Figure 2), our sample was limited to participants in the United States, whose beliefs about the natural and social worlds do not represent the variety across the global population (Heinrich et al., 2010). For example, the specific functions attributed to different social categories likely vary across cultures, as might judgments about who ought to fulfill their functions and how natural they are.

Despite these potential cross-cultural variations, people's views of gender categories likely incorporate both beliefs about natural causes and beliefs about the actions of intentional agents within social structures; these beliefs likely vary even within a single cultural context across different categories and individual people. For these reasons, understanding when and how people think of category-specific behaviors in ways that license normative expectations about what people ought to do (e.g., in terms of functions) is a crucial step towards unravelling rigid systems of gender norms.

References

Archer, J. (1996). Sex differences in social behavior: Are the social role and evolutionary explanations compatible?. *American Psychologist*, *51*(9), 909.

Aristotle (1996). Physics (R. Waterfield & D. Bostock, Trans.). Oxford University Press. (Original work published ca. 350 B.C.).

Barnes, M. E., Evans, E. M., Hazel, A., Brownell, S. E., & Nesse, R. M. (2017). Teleological reasoning, not acceptance of evolution, impacts students' ability to learn natural selection. *Evolution: Education and Outreach*, *10*(1).

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, 67(1), 1–48.

Becker, J. C., & Wright, S. C. (2011). Yet another dark side of chivalry: Benevolent sexism undermines and hostile sexism motivates collective action for social change. *Journal of Personality and Social Psychology*, *101*(1), 62.

Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, *355*(6323), 389-391.

Bian, L., Leslie, S. J., Murphy, M. C., & Cimpian, A. (2018). Messages about brilliance undermine women's interest in educational and professional opportunities. *Journal of Experimental Social Psychology*, *76*, 404-420.

Bloom, P. (1996). Intention, history, and artifact concepts. *Cognition*, *60*(1), 1-29.

Brescoll, V., & LaFrance, M. (2004). The Correlates and Consequences of Newspaper Reports of Research on Sex Differences. *Psychological Science, 15*(8), 515–520.

Buss, D. M. (1995). Evolutionary psychology: A new paradigm for psychological science. *Psychological Inquiry*, *6*(1), 1-30.

Cassidy, B. S., & Krendl, A. C. (2019). A crisis of competence: Benevolent sexism affects evaluations of women's competence. *Sex Roles*, *81*(7), 505-520.

Champely, S. (Developer), Ekstrom, C. (Developer), Dalgaard, P. (Developer), Gill, J. (Developer), Weibelzahl, S. (Developer), Anandkumar, A. (Developer), Ford, C. (Developer), Volcic, R. (Developer), & De Rosario, H. (Developer). (2017). pwr: Basic functions for power analysis. Software https://cran.r-project.org/web/packages/pwr/

Chestnut, E. K., Lei, R. F., Leslie, S. J., & Cimpian, A. (2018). The myth that only brilliant people are good at math and its implications for diversity. *Education Sciences*, *8*(2), 65.

Dar-Nimrod, I., Heine, S. J., Cheung, B. Y., & Schaller, M. (2011). Do scientific theories affect men's evaluations of sex crimes?. *Aggressive Behavior,* 37(5), 440-449.

Del Pinal, G., & Reuter, K. (2017). Dual character concepts in social cognition: Commitments and the normative dimension of conceptual representation. *Cognitive Science*, 41, 477-501.

Diesendruck, G. (2021). A Motivational Perspective on the Development of Social Essentialism. *Current Directions in Psychological Science*, *30*(1), 76-81.

DiYanni, C., & Kelemen, D. (2005). Time to get a new mountain? The role of function in children's conceptions of natural kinds. *Cognition*, *97*(3), 327-335.

Eagly, A. H. (1987). Sex differences in social behavior: A social-role analysis. Hillsdale, NJ: Erlbaum

Eagly, A. H., Wood, W., & Diekman, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. *The Developmental Social Psychology of Gender*, *12*, 174.

Eastwick, P. W., Eagly, A. H., Glick, P., Johannesen-Schmidt, M. C., Fiske, S. T., Blum, A. M., ... & Volpato, C. (2006). Is traditional gender ideology associated with sex-typed mate preferences? A test in nine nations. *Sex Roles*, *54*(9-10).

Fast, A., & Olson, K. R. (2018). Gender development in transgender preschool children. Child Development, 89, 620–637.

Foster-Hanson, E., & Lombrozo, T. (2021, November 24). How "is" shapes "ought" for folk-biological concepts. https://doi.org/10.31234/osf.io/k2sfm

Fox J, Weisberg S (2019). *An R Companion to Applied Regression*, Third edition. Sage, Thousand Oaks CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Gelman, S. A., & Markman, E. M. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development*, 1532-1541.

Gelman, S. A., & Taylor, M. G. (2000). Gender essentialism in cognitive development. In P. H. Miller, & S. E. Kofsky (Eds.), *Toward a feminist developmental psychology*. Florence, KY: Taylor & Frances/Routledge.

Gentner, D., & Kurtz, K. J. (2005). Relational categories. In W.-k. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *Categorization inside and outside the laboratory: Essays in honor of Douglas L. Medin* (pp. 151–175). American Psychological Association.

Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, 70, 491–512.

Glick, P., & Fiske, S. T. (1999). The Ambivalence toward Men Inventory: Differentiating hostile and benevolent beliefs about men. *Psychology of Women Quarterly*, 23, 519 – 536.

Gregory, T. R., & Ellis, C. A. (2009). Conceptions of evolution among science graduate students. *BioScience*, 59(9), 792-799.

Haslanger, S. (2014). The normal, the natural and the good: Generics and ideology. *Politica & Società*, *3*(3), 365-392.

Heider, F. (1958). *The psychology of interpersonal relations.* Wiley, New York.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and Brain Sciences*, *33*(2-3), 61-83.

Ismail, I., Martens, A., Landau, M. J., Greenberg, J., & Weise, D. R. (2012). Exploring the effects of the naturalistic fallacy: Evidence that genetic explanations increase the acceptability of killing and male promiscuity. *Journal of Applied Social Psychology*, 42(3), 735-750.

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology*, 33(1), 1-27.

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, 25(6), 881-919.

Kay, A. C., Gaucher, D., Peach, J. M., Laurin, K., Friesen, J., Zanna, M. P., & Spencer, S. J. (2009). Inequality, discrimination, and the power of the status quo: Direct evidence for a motivation to see the way things are as the way they should be. *Journal of Personality and Social Psychology*, 97(3), 421–434.

Kalish, C. W. (2012). Generalizing norms and preferences within social categories and individuals. *Developmental Psychology*, *48*(4), 1133.

Kalish, C. W., & Lawson, C. A. (2008). Development of social category representations: Early appreciation of roles and deontic relations. *Child Development*, *79*(3), 577-593.

Kelemen, D. (1999a). Beliefs about purpose: On the origins of teleological thought. In M. C. Corballis & S. E. G. Lea (Eds.), *The descent of mind: Psychological perspectives on hominid evolution.* Oxford University Press.

Kelemen, D. (1999c). The scope of teleological thinking in preschool children. *Cognition*, *70*(3), 241–272.

Kelemen, D. (2004). Are children "intuitive theists"? Reasoning about purpose and design in nature. *Psychological Science*, *15*(5), 295–301.

Kelemen, D., & Diyanni, C. (2005). Intuitions about origins: Purpose and intelligent design in children's reasoning about nature. *Journal of Cognition and Development*, *6*(1), 3–31.

Kelemen, D., & Rosset, E. (2009). The Human function compunction: Teleological explanation in adults. *Cognition*, *111*(1), 138–143.

Kelemen, D., Rottman, J., & Seston, R. (2013). Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default. *Journal of Experimental Psychology: General*, *142*(4), 1074–1083.

Knobe, J., Prasada, S., & Newman, G. E. (2013). Dual character concepts and the normative dimension of conceptual representation. *Cognition*, 127(2), 242-257.

Kray, L.J., Howland, L., Russell, A.G., & Jackman, L.M. The effects of implicit gender role theories on gender system justification: Fixed beliefs strengthen masculinity to preserve the status quo. *Journal of Personality and Social Psychology*. 2017; 112(1):98–115.

Kuznetsova, A., Brockhoff, P.B., & Christensen, R.H.B. (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software*, 82(13), 1–26.

Lane, M. (2020, September 9). Normative theory and the perils of normalization. [Paper for panel on "Things as They Should Be?" An Interdisciplinary Conversation in the Humanities]. 2020 Humanities Colloquium, Princeton, NJ, United States.

Lewry, C., Kelemen, D., & Lombrozo, T. (2021). Why belief in species purpose prompts moral condemnation of individuals who fail to fulfill that purpose. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43, No. 43).

Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition*, 99(2), 167-204.

Lombrozo, T., & Rehder, B. (2012). Functions in biological kind classification. *Cognitive Psychology*, *65*(4), 457-485.

Lombrozo, T., Shtulman, A., & Weisberg, M. (2006). The Intelligent Design Controversy: Lesson From Psychology and Education. *Trends in Cognitive Sciences*, 10(2), 56.

Malle, B. F. (2011). Attribution theories: How people make sense of behavior. *Theories in Social Psychology*, *23*, 72-95.

Mayr, E. (1982). The growth of biological thought: Diversity, evolution, and inheritance. Cambridge, MA: Harvard University Press.

Meyer, M., Cimpian, A., & Leslie, S. J. (2015). Women are underrepresented in fields where success is believed to require brilliance. *Frontiers in Psychology*, *6*, 235.

Meyer, M., & Gelman, S. A. (2016). Gender essentialism in children and parents: Implications for the development of gender stereotyping and gender-typed preferences. *Sex Roles*, *75*(9), 409-421.

Rips, L. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning*. Cambridge: Cambridge University Press.

Roberts, A. J., Handley, S. J., & Polito, V. (2021). The design stance, intentional stance, and teleological beliefs about biological and non-biological natural entites. *Journal of Personality and Social Psychology*, *120*(6), 1720–1748.

Rose, D., & Nichols, S. (2019). Teleological essentialism. *Cognitive Science*, *43*(4), e12725.

Rose, D., & Nichols, S. (2020). Teleological essentialism: generalized. *Cognitive Science*, *44*(3), e12818.

Shtulman, A. (2006). Qualitative differences between naïve and scientific theories of evolution. *Cognitive Psychology*, 52(2), 170-194.

Shtulman, A. (2017). Scienceblind: why our intuitive theories about the world are so often wrong. New York: Basic Books.

Skewes, L., Fine, C., & Haslam, N. (2018). Beyond Mars and Venus: The role of gender essentialism in support for gender inequality and backlash. *PloS one*, *13*(7), e0200921.

Springer, K., & Keil, F. C. (1989). On the development of biologically specific beliefs: The case of inheritance. *Child Development*, 637-648.

Strohminger, N., Knobe, J., & Newman, G. (2017). The true self: A psychological concept distinct from the

self. *Perspectives on Psychological Science*, *12*(4), 551-560.

Tooby J, Cosmides L. 1992. The psychological foundations of culture. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, ed. JH Barkow, L Cosmides, J Tooby. New York: Oxford Univ. Press

Ware, E. A., & Gelman, S. A. (2014). You get what you need: An examination of purpose-based inheritance reasoning in undergraduates, preschoolers, and biological experts. *Cognitive Science*, 38(2), 197-243.