



## Gender Categories as Dual-Character Concepts?

Cai Guo,  Carol S. Dweck, Ellen M. Markman

*Department of Psychology, Stanford University*

Received 28 March 2020; received in revised form 29 January 2021; accepted 1 February 2021

---

### Abstract

Seminal work by Knobe, Prasada, and Newman (2013) distinguished a set of concepts, which they named “dual-character concepts.” Unlike traditional concepts, they require two distinct criteria for determining category membership. For example, the prototypical dual-character concept “artist” has both a concrete dimension of artistic skills, and an abstract dimension of aesthetic sensibility and values. Therefore, someone can be a good artist on the concrete dimension but not truly an artist on the abstract dimension. Does this analysis capture people’s understanding of cornerstone social categories, such as gender, around which society and everyday life have traditionally been organized? Gender, too, may be conceived as having not only a concrete dimension but also a distinct dimension of abstract norms and values. As with dual-character concepts, violations of abstract norms and values may result in someone being judged as not truly a man/woman. Here, we provide the first empirical assessment of applying the dual-character framework to people’s conception of gender. We found that, on some measures that primarily relied on metalinguistic cues, gender concepts did indeed resemble dual-character concepts. However, on other measures that depicted transgressions of traditional gender norms, neither “man” nor “woman” appeared dual-character-like, in that participants did not disqualify people from being truly a man or truly a woman. In a series of follow-up studies, we examined whether moral norms have come to replace gender role norms for the abstract dimension. Implications for the evolution of concepts and categories are explored.

*Keywords:* Concepts; Normative thinking; Gender; Morality; Folk theory

---

---

Correspondence should be sent to Cai Guo, Department of Psychology, Stanford University, Building 420, 450 Jane Stanford Way, Stanford, CA 94305. E-mail: caigu@stanford.edu

A portion of the results of the current project was previously reported in a conference proceeding of the 40th Annual Meeting of the Cognitive Science Society, Madison, WI, 2018. The current submission includes substantial extensions on the results reported in the conference proceeding as well as important new experiments not reported in the conference proceeding.

## 1. Introduction

### 1.1. Overview

The psychological basis for categorization has been a central topic in cognitive science for decades. What makes something a member of a category? Starting from Wittgenstein's (1953) foundational insights about family resemblance, some of the most important findings in cognitive science grow out of these questions, such as the probabilistic versus deterministic nature of categorization (Rosch & Mervis, 1975) and the roles of category ideals versus prototypes in evaluating category members (Barsalou, 1985; Lakoff, 1987).

Even after decades of transformative research, the field is still witnessing new insights and answers to these questions. One recent advance is the *dual-character concepts framework* proposed by Knobe, Prasada, and Newman (2013), which posits that some categories require evaluation on two separate dimensions (a concrete and an abstract dimension) to assess membership. In this paper, we examine the extent to which the dual-character concept framework can advance our understanding of a cornerstone social concept, namely, gender. We focus on gender because it is an extremely important social category and also because there remain some interesting inconsistencies in the literature on people's conceptualization of gender categories that the dual-character concept framework may shed light on. For example, ordinary expressions such as "real men don't eat quiche" or "real men don't cry" are readily comprehensible, yet they are simultaneously saying that a given man is a man but not a "real man." As Leslie (2015) argued, gender categories may function like dual-character concepts, as they may encompass two separate, independent dimensions that can each lead to distinct category membership judgments at the same time. We empirically examine this possibility and further discuss how such an examination not only advances our understanding of gender concepts but also provides insights into critical aspects of the dual-character framework and conceptual representations in general.

### 1.2. Dual-character concepts

*Dual-character concepts* are a distinct set of concepts that involve two separate, independent conceptual dimensions for evaluating category membership, one of which concerns more concrete traits and features and the other of which concerns abstract norms and attributes (Knobe et al., 2013). For example, the prototypical dual-character concept "artist" involves a concrete dimension of having appropriate technical skills and an abstract dimension of fulfilling the norms and values that we typically attribute to artists (e.g., demonstrating exceptional artistic sensibility and creating original work that embodies aesthetic values).

Only concepts to which we typically attribute abstract values in addition to the concrete features can be candidates for dual-character concepts. According to Knobe et al., one way to examine whether a concept has abstract values is the perceived naturalness of describing the concept with both "good" and "true," with each adjective indicating a

separate dimension. For instance, for dual-character concepts like “artist,” it should sound natural to say “a good artist,” representing someone who has the technical abilities as an artist as well as to say “a true artist,” representing someone who embodies the more abstract values, such as aesthetic values and a commitment to maintaining these values (Del Pinal & Reuter, 2017). In contrast, for concepts such as “bus driver” to which we do not typically attribute abstract values, it should only sound natural to most people to say “a good bus driver” but not “a true bus driver.”

Knobe et al. also showed that for dual-character concepts, judgments about category membership on one dimension can be distinct from judgments on the other. Specifically, they found that for dual-character concepts such as “artist,” participants found it natural to say that someone is clearly an artist in one sense but ultimately is not truly an artist in another sense. To flesh this out, Knobe et al. had participants read vignettes about specific individuals violating abstract norms. They found that for dual-character concepts, such as “artist,” participants would agree, for example, that someone who has sufficient technical skills but creates art only for money and does not care about originality or the aesthetic values is an artist in one sense but not truly an artist in another sense.

Notably, as Leslie (2015) pointed out, the abstract dimension of dual-character concepts is not simply raising the standard for the concrete dimension. For example, even though extraordinary or exemplary bus drivers exist, this does not render “bus driver” a dual-character concept. That is, unlike dual-character concepts such as “artist,” a bus driver who cares greatly about punctuality and passengers’ safety may be viewed as an admirable, responsible bus driver, but as Knobe et al. (2013) showed, participants did not think that the lack of these admirable traits would make someone not truly a bus driver. Instead, for such concepts, participants thought that a bus driver who is brusque and does not care about the passengers would still be a full-fledged bus driver. That is, these admirable or negative qualities are being evaluated within the concrete dimension and are not operating as a distinct normative dimension.

Moreover, Knobe et al. (2013) showed that for natural kind concepts, for which category membership is often determined by beliefs about an invisible “essence” (Gelman, 2003; Keil, 1989; Malt, 1990; Medin & Ortony, 1989), people did not express two distinct judgments either (see Tobia, Newman, & Knobe, 2019, for water as an exception). In short, unlike dual-character concepts, neither concepts like “bus driver” nor natural kind concepts like “tiger” showed such a dual-character structure where category membership can be judged on the concrete and abstract dimensions separately.

The converging findings across a variety of tests in Knobe et al. (2013) provided strong support for the unique features of dual-character concepts and established a systematic approach for examining and broadening prior findings and insights from the field. The early major advances in theories about categorization, such as family-resemblance-based (Rosch & Mervis, 1975; Wittgenstein, 1953) and exemplar-based categorization (Medin & Schaffer, 1978), mostly emphasized the nature of categorization as probabilistic rather than defined by necessary and sufficient features. These probabilistic views, however, still posit categorization as forming a single gradient based either on the category prototype or on reference to a single set of previous exemplars.

Subsequent research, however, further refined these seminal views and paved the way for dual-character concepts. Barsalou (1985), for example, showed that category membership might be influenced not only by the gradient of similarity to the average exemplar or prototype of the category (central tendency) but also by how close the instance is to the ideals of the category (Barsalou, 1985). This contrast between ideals and central tendency led to further speculation that for some categories, what enables an instance to qualify as a member of a category in one context might be different and distinct from the enabling criteria in other contexts, which is a direct antecedent to the dual-character framework. Lakoff (1987) in *Women, Fire, and Dangerous Things* provided pioneering insights in this regard. He suggested, for example, that the concept “mother” might have multiple models for determining category membership, such that a “real” mother can be a biological mother, a mother who raises the child, a mother with legal custody of the child, and so on. Judgments about being a “real” mother, therefore, may be based on an entirely different set of criteria in one context than in another. Building on such insights, the dual-character framework goes a step further to hypothesize that the different models for concepts like “mother” can be systemized onto two separate dimensions, one concrete (such as biological relationship) and one abstract (such as fulfilling a mother’s role), each of which can grant or deny category membership separately. Another potential contribution of this framework is its ability to demonstrate how the two dimensions can operate simultaneously yet separately, such that a person can be viewed as clearly a mother in the concrete sense of biology but, at the same time, not truly a mother in the sense of providing nurturance and fulfilling a mother’s role. The dual-character framework may also bear on Putnam (1973)’s famous “Twin Earth” scenario involving a clear liquid on the “Twin Earth” that looks and tastes just like water but is not composed of H<sub>2</sub>O molecules. Putnam believed that the “Twin Earth” liquid is not water because it is the chemical component H<sub>2</sub>O, rather than taste and appearance, that defines “water.” Challenging such purely essentialist view, Malt (1994) demonstrated that, despite people’s knowledge that “water” is H<sub>2</sub>O, they still granted “water” status to liquids whose primary component was not H<sub>2</sub>O if the liquids had water-like attributes regarding their source, location, and function. In a recent study, Tobia et al. (2019) expanded on Malt’s findings arguing that “water” is a dual-character concept that is understood by means of two separate dimensions that could each grant (or deny) category membership separately.

The method used for detecting possible dual-character concepts is partly inspired by research in psycholinguistics on linguistic hedges, since hedges serve as cues to focus on technical definitions or more abstract traits in judging category membership (Lakoff, 1973, 1987; Malt, 1990). As noted earlier, a key signature of a dual-character concept is how felicitous it is to describe a concept with “true,” such that only for dual-character concepts is “true” perceived as natural, whereas for concepts like “bus driver,” it sounds inappropriate to describe them with “true.” The use of “true” as a signature of dual-character concepts echoes previous research by Lakoff (1973, 1987) that provided the seminal intuition that some linguistic hedges in everyday language, such as “technically” and “strictly speaking,” signal that different features can be emphasized in category membership judgments. For instance, saying “technically Richard Nixon is a Quaker”

highlights the concrete features of the “Quaker” category that Nixon possessed, such as being an official member of the religious group, as opposed to the more abstract values, such as holding pacifist views.

Moreover, the application of “true” in dual-character concepts framework is inspired by but also different from “true” judgments in the judgments about authenticity or originality. Unlike judgments such as “a true Picasso painting” (Newman & Bloom, 2012), “true” in the case of dual-character concepts is not a marker of being authentic versus fake. Instead, the “true” hedge in the dual-character framework implies the existence and salience of the abstract dimension for a concept and thus the possibility of the concept being a dual-character concept.

In short, the dual-character concept framework helps to clarify and reconcile the previously identified inconsistencies and contradictions in category membership judgments by postulating a distinct kind of concept for which fulfillment of abstract values is important and independent of concrete, technical features in determining category membership.

### *1.3. Gender categories as dual-character concepts*

Since its introduction, the dual-character framework has gained considerable attention in the field and has inspired discussions on the potential of the framework for illuminating psychological representations in social domains (Reuter, 2019), but so far there has been little research on how the framework may advance our understanding of cornerstone social concepts such as gender, around which society and everyday life have traditionally been organized. We chose to focus on gender in the present research also because, as Leslie (2015) noted, the dual-character concept framework might shed light on the intriguing inconsistencies in the folk psychology of gender. Specifically, in the gender literature, there remains a contradiction in how people mentally represent gender as a category (Leslie, 2015; Smiler & Gelman, 2008; Vandello, Bosson, Cohen, Burnaford, & Weaver, 2008). On the one hand, it is well documented that by the early school years, children have a full grasp of gender constancy, a developmental hallmark reflecting the awareness that simple transgressions of gender norms do not change a person’s gender (Kohlberg, 1966; Ruble et al., 2007). It is also well documented that people often endorse gender essentialism which views gender as immutable and biologically determined (Gelman, 2003; Prentice & Miller, 2007).

On the other hand, as social psychologists have found, a person’s gender can also be viewed as “precarious,” such that one’s gender membership may be questioned or disqualified if one does not actively “prove” one’s gender by showing gender normative behavior (Vandello et al., 2008). Past research suggested that there have been rules that one must follow and behavioral traits one must exhibit to qualify as a “real man” (e.g., being tough, being a financial provider, and being protective of family) or “real woman” (e.g., being nurturing, gentle, and communal) or otherwise risk disqualifying oneself as a real or true member of the gender category (Eagly & Karau, 2002; Gilmore, 1990; Leslie, 2015; Vandello & Bosson, 2013; Vandello et al., 2008). This inconsistency has puzzled social, developmental, and cognitive psychologists alike (Eidson & Coley, 2014; Vandello et al., 2008).

Leslie (2015) suggests that “man” and “woman” might very well be dual-character concepts such that there are two independent, parallel dimensions corresponding to two different senses of gender concepts—a concrete dimension of gender identity, and an abstract dimension of gender norms and values, such as being dominant for men and nurturing for women. Judgments about gender membership on the two dimensions are therefore separable and can be different from each other. Leslie’s (2015) intuitions seem to resonate with recent research showing that gender essentialism only predicts judgments about more concrete features but not more abstract behavioral properties (Eidson & Coley, 2014). Moreover, the broad gender literature has increasingly endorsed a multidimensional construal of gender, such that “gender” is an umbrella entity consisting of separate constructs, including “being gender,” which concerns an individual’s current gender identity, and “doing gender,” which concerns such aspects as gender roles and norms and gender social presentations (Hyde, Bigler, Joel, Tate, & van Anders, 2019; Tate, Youssef, & Bettergarcia, 2014). It is therefore possible that, as Leslie suggested, the belief that gender identity is constant concerns the concrete dimension of gender concepts, which is separate from the abstract dimension of values and norms where gender membership may be precarious.

Given the findings in Knobe et al. (2013) and given Leslie’s (2015) philosophical analysis, gender concepts may indeed have the potential to function in the same ways as the dual-character concepts in Knobe et al. (2013). Empirical examination of gender as a dual-character concept, however, is currently lacking. Such examination is important, as it might not only help us understand the paradox in the folk psychology of gender that has long been awaiting an answer but may also shed light on some critical aspects of the dual-character framework itself and conceptual representations in general. Much is still unknown about what kinds of concepts can be dual-character concepts and how we determine what features are central to category membership for each dimension, especially the abstract, normative dimension. The sample dual-character concepts examined in Knobe et al. (2013) are all quite specific social roles (e.g., artist) or specific conceptual constructs with social meanings (e.g., poetry and love). However, unlike “artist” or “poetry,” there are other much more socially significant concepts, such as gender, around which society has historically been organized and in terms of which virtually all people, not just subsets of them, have been understood (Maccoby, 1998). Consequently, examining gender categories under the dual-character framework is a natural further step for assessing the framework’s utility in understanding key social concepts, and can also provide a test of the boundaries of dual-character concepts. In this paper, we present the first empirical examination of whether gender categories are indeed represented as dual-character concepts in people’s lay beliefs.

#### *1.4. The current design*

In the current research, the first four experiments were straightforward replications of the work by Knobe et al. (2013), with the addition of the two gender categories “man” and “woman.” The replications sought to examine whether gender concepts indeed

function as dual-character concepts based on Knobe et al.'s original criteria. The next three experiments addressed the surprising findings that emerged from the first four experiments, including why the dual-character signature for gender was stronger on some tasks than others, and why “man” showed greater consistency than “woman” in demonstrating the dual-character signature.

In summary, the current work sought to use the dual-character concept framework to shed light on the paradox inherent in folk theories of gender concepts and, in doing so, to examine, and perhaps refine the nature of dual-concept categories.

## 2. Experiment 1

Our first experiment was a direct replication of Experiment 1 in Knobe et al. (2013) with the addition of two gender concepts. Specifically, Knobe et al. argued that one way of testing dual-character concepts is to see whether the concepts can be naturally described with both “good” and “true” adjectives. Knobe et al. found that when participants judged whether statements sounded natural, dual-character concepts were rated significantly higher than were other concepts when described with the “true” adjective in the statements (e.g., she’s a true artist vs. he’s a true cashier), although not when they were described with the “good” adjective (e.g., she’s a good artist vs. he’s a good cashier). Following Knobe et al., the current experiment aimed to assess whether gender concepts resemble dual-character concepts in the extent to which they can be described with “true” (a true man), which concerns abstract traits, as well as “good” (a good man), which concerns concrete traits.

### 2.1. Method

#### 2.1.1. Participants

We recruited 161 participants on Amazon Mechanical Turk ( $M_{\text{age}} = 32.48$ ; age range: 19–68; 38.5% female; 60.9% male; 0.6% nonbinary). See below for information about power analysis.

#### 2.1.2. Materials and procedure

Participants completed an online survey with 42 pairs of statements (20 dual-character, 20 control, and 2 gender concepts) in a randomized order. All the statements, except those for gender concepts, were directly adopted from Experiment 1 in Knobe et al. (2013). Each pair of statements for a specific concept contained a “good” statement (e.g., Marie is a good artist) and a “true” statement (e.g., Marie is a true artist) that indicated possession of concrete traits and abstract values, respectively. Participants rated both types of statements on a 7-point Likert scale, ranging from 1(*sounds weird*) to 7(*sounds natural*).

## 2.2. Data analysis plan

### 2.2.1. Linear mixed-effects regression models

For all the experiments that we reported in this paper, we first conducted null hypothesis testing analyses by fitting linear mixed-effects regression (LMER) models in R with Satterthwaite approximations for degrees of freedom and a maximal random-effects structure as recommended by Barr, Levy, Scheepers, and Tily (2013) and Bates, Kliegl, Vasishth, and Baayen (2015), which is robust against Type I error, including in multiple comparisons (Gelman, Hill, & Yajima, 2012). For by-item and by-subject random effects, we included both random slopes and intercepts for effects that were both within- and between-subjects (items) and included random intercepts for effects that were only between-subjects (items). An R example of the type of LMER model with maximal random-effects structure for the interaction between concept and statement types would be `lmer(value ~ Concept_Type × Statement + (1 + Statement|Item) + (1 + Concept_Type × Statement|Subject))`. Whenever the maximal random-effects models did not converge, we followed Bates et al., (2015)'s suggestion to simplify the models until they successfully converged (see <https://osf.io/f5y2q/> for specific model parameterization and all codes for analyses).

For all experiments conducted, we first compared whether “man” and “woman” were significantly different from each other. If “man” and “woman” were not different on the relevant measures, we would combine the two types into a single concept type “gender” in subsequent analyses that compared gender concepts to dual-character and other types of concepts.

### 2.2.2. Bayesian analysis using Bayes factors

Fitting maximal LMER models for extremely unbalanced groups would likely lead to reduced power (Barr et al., 2013; Eager & Roy, 2017). In our case, there was only one pair of items for each gender concept but 20 pairs of items for each of all the other concept types. Therefore, the maximal LMER models could be excessively conservative for the current unbalanced design, leading to a higher Type II error rate by penalizing comparisons between the unbalanced groups.

To solve this issue, we turned to Bayesian model comparisons, conducted with the BayesFactor package in R (Morey, Rouder, & Jamil, 2015), which also considers the random effects of specific concept items and each individual participant (Rouder & Morey, 2012) but is much less susceptible to the unbalanced design problem (Bayman, Chaloner, & Cowles, 2010; Kruschke, 2015). Bayes factors also obviate the concern about Type I error, and moreover, can directly demonstrate which hypothesis has the most evidence given the data collected, and the amount of evidence for each alternative hypothesis can be directly compared. Therefore, for all the experiments that involved an unbalanced design in the current study, we will report results from both the LMER models and the Bayes factor comparisons (see Section 1.1 in Appendix S1 for the specifics of the Bayesian approach).



Given the difficulty of estimating sample sizes from the unbalanced design, we included samples that were approximately four to five times larger than the original sample sizes in Knobe et al. (2013).

### 2.3. Results and discussion

#### 2.3.1. LMER models

We first compared whether “man” and “woman” differed significantly from each other in participants’ ratings for the “good” and “true” statements. The results revealed no difference between “man” and “woman,” as there was neither a significant interaction between concept (“man” vs. “woman”) and statement type (“good” vs. “true”),  $b = 0.24$ ,  $SE = 0.16$ ,  $t(320) = 1.51$ ,  $p = .13$ , nor a difference in overall ratings for “man” and “woman” across statement types,  $b = 0.08$ ,  $SE = 0.08$ ,  $t(320) = 0.97$ ,  $p = .33$ . Thus, we treated them as the single category “gender” in subsequent analyses.

We then examined whether there were significant differences among dual-character, control, and gender concepts in how natural participants perceived the “good” and “true” statements to be for each kind of concepts. The results revealed a significant interaction between concept and statement types,  $F(2, 39) = 11.21$ ,  $p < .001$ . The interaction was driven by the fact that while the three types of concepts did not differ significantly from one another in participants’ ratings for “good” statements, ( $bs < 0.31$ ,  $ts < 0.61$ ,  $ps > .54$ ), they differed significantly, as predicted, for “true” statements, such that both dual-character and gender concepts had higher ratings in “true” statements than did control concepts ( $bs > 1.34$ ,  $ts > 2.63$ ,  $ps < .012$ ), and gender concepts did not differ from dual-character concepts in this regard,  $b = -0.35$ ,  $SE = 0.51$ ,  $t(39.2) = -0.69$ ,  $p = .50$  (see Fig. 1).

In addition to examining how the three concept types differed in “good” and “true” statements separately, we compared the concept types against one another to determine whether the difference between ratings for “good” statements and ratings for “true statements” varied for dual-character, control, and gender concepts. The results revealed that regarding the difference between ratings for “good” and “true” statements, gender concepts were *not* different from either control,  $b = 1.04$ ,  $SE = 0.83$ ,  $t(39) = 1.24$ ,  $p = .22$ , or dual-character concepts,  $b = -0.65$ ,  $SE = 0.83$ ,  $t(39) = 0.78$ ,  $p = .44$ . Dual-character concepts were, however, significantly different from control concepts in this regard,  $b = 1.68$ ,  $SE = 0.36$ ,  $t(39) = 4.73$ ,  $p < .001$ . As previously mentioned in the data analysis plan, the fact that gender concepts were not different from either dual-character or control concepts in this regard but the latter two in fact were significantly different from each other might partly result from the unbalanced design. The standard error was much smaller ( $SE = 0.36$ ) in the comparison between the two equal-sized types, dual-character and control concepts, than in the unbalanced comparisons between gender and the other two types ( $SEs = 0.83$ ). We therefore turned to Bayesian model comparison approach using Bayes factor for a more accurate assessment.

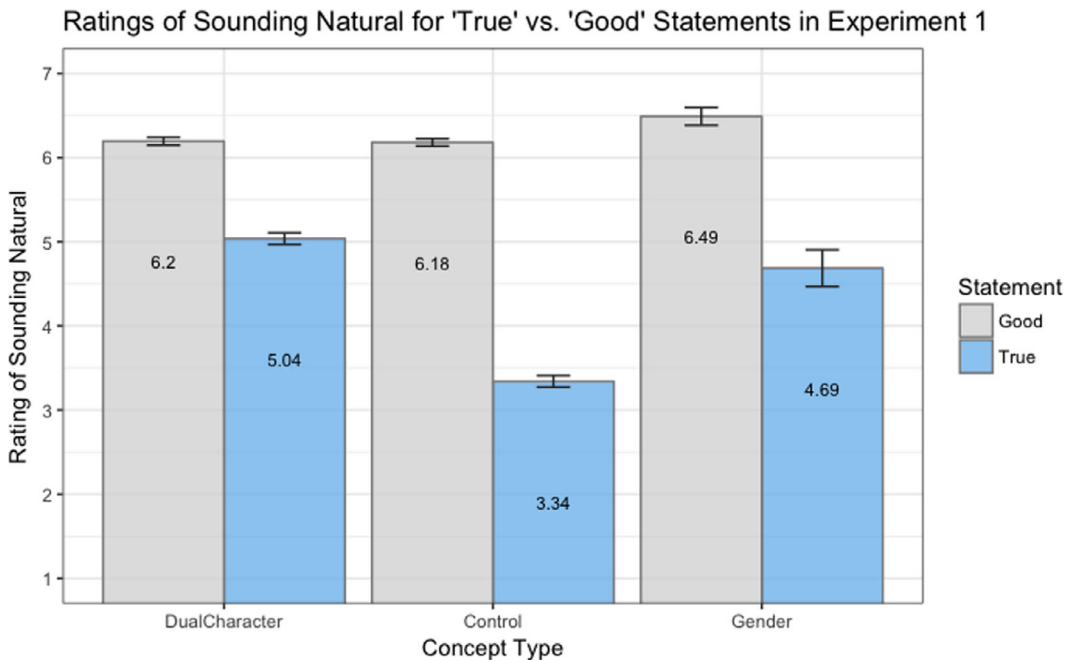


Fig. 1. Participants' ratings of how natural "true" and "good" statements sounded for dual-character, control, and gender concepts in Experiment 1. Error bars represent bootstrapped 95% confidence intervals.

### 2.3.2. Bayes factors

The Bayes factors obtained for all the possible hypotheses regarding how the concept types compared to one another showed that the model that treated each concept type independently yielded the largest Bayes factor (see Table A in Section 2.1 in Appendix S1). However, the model that equated "gender" with dual-character concepts was the next best model—when we directly compared the all-different model against the "gender = dual-character" model, the former outperformed the latter by a relatively short margin (a Bayes factor of 254). More importantly, when we directly compared the "gender = dual-character" against the "gender = control" model, the former outperformed the latter by a large margin (a Bayes factor of  $3.23 \times 10^{37}$ ). Therefore, gender concepts were indeed much closer to dual-character concepts than to control concepts in this regard.

## 2.4. Conclusion

In conclusion, the current experiment replicated Knobe et al.'s results such that dual-character concepts did not differ from control concepts in "good" statements but had considerably higher ratings than control concepts in "true" statements. Overall, just like dual-character concepts, the gender concepts "man" and "woman" seemed to also possess a conceptual dimension of abstract ideals, as evidenced by the finding that describing a man or woman as a "true man" or "true woman" was perceived to be felicitous.

### 3. Experiment 2

In Experiment 2, we turned to replicating the next task in Knobe et al. (2013), one which they felt could provide further support for the independence of the concrete and abstract conceptual dimensions of dual-character concepts. Specifically, Knobe et al. (2013) assessed whether participants would consider it natural to say that someone is a member of a category in one sense but not the other. They found that only for dual-character concepts would participants consider both statements to sound natural. For example, participants thought that it would sound natural to say that there is a sense in which Amy is an artist but there is a sense in which Amy is *not* truly an artist. Similarly, it would also sound natural to say that there is a sense in which Amy is *not* an artist but there is a sense in which Amy is truly an artist after all. In comparison, for concepts such as “cashier,” neither statement would sound as natural as for dual-character concepts.

This result thus suggested that only for dual-character concepts, but not control concepts, would participants make two separate judgments at the same time about a person’s category membership; that is, the person could be granted membership on the concrete dimension but be disqualified of membership on the abstract dimension or vice versa. To examine whether gender concepts would also resemble dual-character concepts in this regard, we conducted Experiment 2 following the same procedures in Knobe et al., with the addition of “man” and “woman.”

#### 3.1. Method

##### 3.1.1. Participants

We recruited 150 participants on Amazon Mechanical Turk. Three participants failed to pass the attention check question, leading to a final sample of 147 participants ( $M_{\text{age}} = 34.82$ ; age range: 18–67; 49.7% female; 50.3% male).

##### 3.1.2. Materials and procedure

In their original Experiment 2, Knobe et al. (2013) included only the 10 most dual-character-like and the 10 most control-like concepts as revealed in their Experiment 1. That is, in Knobe et al.’s Study 1, there was a range of average agreement about the extent to which a concept was dual-character-like or control-like, and only the clearest cases were chosen for further study. To replicate Knobe et al., we included only the same 10 most dual-character-like concepts (artist, criminal, mentor, love, friend, comedian, minister, boyfriend, theory, and argument) and the same 10 most control-like concepts (rustling, welder, catalog, chair, firefighter, uncle, cashier, stroller, obituary, second cousin) as in their original study, rather than the full set from Experiment 1, with the addition of “man” and “woman.” For each specific concept, two statements were constructed: (a) “There’s a sense in which X is clearly a . . . , but ultimately, if you think about what it really means to be a X, you’d have to say that there is a sense in which X is not truly a . . . at all” (ultimate nonmember statement); and (b) “There’s a sense in which X is

clearly not a . . . , but ultimately, if you think about what it really means to be a . . . , you'd have to say that there is a sense in which X is truly a . . . after all" (ultimate member statement). Participants received the pairs of statements for each concept in a randomized order, and for each concept, participants rated both statements on a 7-point Likert scale ranging from 1 (*sounds weird*) to 7 (*sounds natural*).

### 3.2. Results and discussion

We first compared "man" and "woman" and found that "man" and "woman" differed significantly from each other: overall, across the two statement types, "man" had higher ratings than did "woman,"  $b = 0.48$ ,  $t(146) = 4.60$ ,  $SE = 0.10$ ,  $p < .001$ . This means that in general the two statements that each granted category membership in one sense and denied it in the other sense were perceived to sound more natural for "man" than for "woman." There was also a significant interaction between concept ("man" vs. "woman") and statement type,  $b = -0.39$ ,  $t(146) = -2.21$ ,  $SE = 0.18$ ,  $p = .03$ , indicating that the difference between participants' ratings for the two statements differed for "man" and "woman," such that participants did not rate the two statements differently for "woman,"  $b < 0.001$ ,  $t(240.6) = 0.00$ ,  $SE = 0.17$ ,  $p = 1.00$ , but rated the "ultimately not a member" statement higher than the "ultimately a member statement" for "man,"  $b = 0.39$ ,  $t(240.6) = 2.29$ ,  $SE = 0.17$ ,  $p = .02$ . Therefore, we separated "man" and "woman" in the subsequent analyses.

We then examined whether dual-character, control, and each of the two gender concepts differed significantly from one another (see Fig. 2) and found that "man" was considerably similar to dual-character concepts but different from control concepts, whereas "woman" was significantly similar to control concepts but different from dual-character concepts. Specifically, the results first revealed a significant interaction between concept type and statement type,  $F(3, 19.25) = 3.14$ ,  $p = .05$ , such that the difference between the participants' ratings for the two types of statements varied for dual-character concepts, control concepts, man, and woman.

To interpret the interaction, we first compared dual-character concepts to control concepts. We found that the results in Knobe et al. (2013) were successfully replicated, such that there was both a significant main effect of concept type (i.e., dual-character vs. control), as dual-character concepts had overall higher ratings than control concepts across statement types (i.e., member and nonmember statements),  $b = 1.24$ ,  $SE = 0.15$ ,  $t(28.4) = 8.48$ ,  $p < .001$ , and a significant interaction between concept type (dual-character vs. control) and statement type,  $b = -0.35$ ,  $SE = 0.14$ ,  $t(19.44) = -2.55$ ,  $p = .02$ . Specifically, the interaction was driven by the fact that there was no difference between member and nonmember statements for dual-character concepts,  $b = 0.02$ ,  $SE = 0.12$ ,  $t(43.4) = -0.15$ ,  $p = .88$ , whereas for control concepts, member statements had significantly higher ratings than did nonmember statements,  $b = -0.33$ ,  $SE = 0.13$ ,  $t(49.7) = 2.62$ ,  $p = .01$ .

We then compared gender concepts to dual-character and control concepts, respectively, to examine whether gender concepts resembled dual-character or control concepts

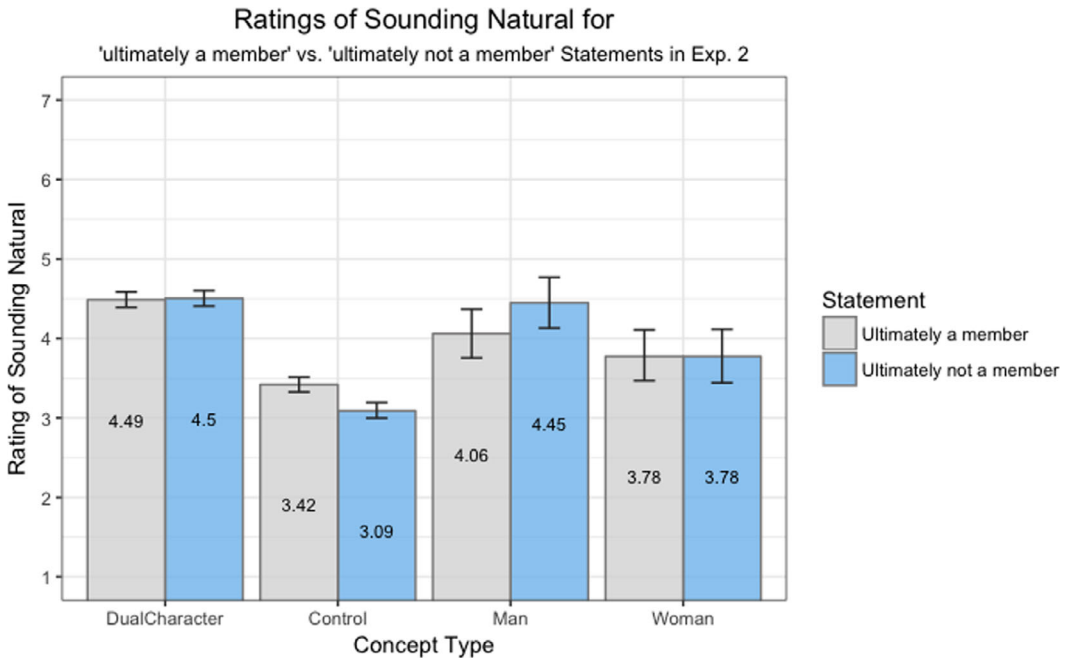


Fig. 2. Participants' ratings of the "ultimately a member" versus the "ultimately not a member" statements for dual-character concepts, control concepts, man, and woman in Experiment 2. Error bars represent bootstrapped 95% confidence intervals.

in this regard. The results showed that "man" was very similar to dual-character concepts, as there was neither a main effect of how "man" and dual-character concepts were rated overall across the statements,  $b = -0.24$ ,  $SE = 0.31$ ,  $t(19.7) = -0.77$ ,  $p = .45$ , nor an interaction between concept type ("man" vs. dual-character) and statement type (ultimate member vs. ultimate nonmember),  $b = 0.37$ ,  $SE = 0.32$ ,  $t(19.00) = 1.17$ ,  $p = .26$ . In contrast, "man" differed significantly from control concepts, such that there was both a significant main effect of concept type ("man" vs. control), as overall across statement types, "man" had significantly higher ratings than did control concepts,  $b = 1.00$ ,  $SE = 0.31$ ,  $t(20.2) = 3.18$ ,  $p = .005$ , and a significant interaction between concept type and statement type,  $b = 0.72$ ,  $SE = 0.32$ ,  $t(19.72) = 2.24$ ,  $p = .037$ . The interaction was driven by the fact that for "man," as previously reported, the nonmember statement had significantly higher ratings than did the member statement, whereas for control concepts, the nonmember statement had significantly lower ratings than did the member statement.

We then compared "woman" and dual-character concepts, and the results showed that although there was no significant interaction between concept type and statement type,  $b = -0.02$ ,  $SE = 0.32$ ,  $t(18.9) = 0.06$ ,  $p = .96$ , there was a significant main effect such that "woman" overall had significantly lower ratings than did dual-character concepts,  $b = -0.72$ ,  $SE = 0.32$ ,  $t(21.1) = 2.27$ ,  $p = .03$ . Similarly, we compared "woman" and control concepts, and the results showed that there was neither a significant main effect

of concept,  $b = 0.52$ ,  $SE = 0.32$ ,  $t(20.7) = 1.65$ ,  $p = .11$ , nor a significant interaction between concept and statement type,  $b = -0.33$ ,  $SE = 0.32$ ,  $t(19.6) = -1.03$ ,  $p = .32$ . Therefore, it seems that “woman” may be closer to control concepts than to dual-character concepts.

As before, to further investigate the distance between gender concepts and dual-character and control concepts, we conducted Bayesian model comparisons and directly compared the Bayes factors of the models (see Table B in Section 3.1 in Appendix S1). The model comparisons showed that the best model was the one that equated “man” with dual-character concepts, which in fact outperformed the all-different model by a small margin (a Bayes factor of 1.34 in favor of the “man = dual-character” model). In addition, the model that equated “man” with control concepts revealed much less evidence than did the model that equated “man” and dual-character concepts (by a Bayes factor of  $3.97 \times 10^{16}$ ), and also less than the all-different model (by a Bayes factor of  $2.96 \times 10^{16}$ ). Therefore, consistent with the results from the LMER models, “man” was much closer to dual-character concepts than to control concepts in this case. In contrast, the model that equated “woman” with dual-character concepts was much less supported by the data than were the all-different model (by a Bayes factor of  $6.73 \times 10^7$ ) and the model that equated “woman” with control concepts (by a Bayes factor of 3,951). Thus, also consistent with the results from the linear-mixed effect models, “woman” seemed closer to control concepts than to dual-character concepts in this study.

### 3.3. Conclusion

In summary, the measure in Experiment 2, which assessed whether the two conceptual dimensions could indeed function independently in category membership judgments, revealed that “man” resembled dual-character concepts in demonstrating independence of the concrete and abstract dimensions. Unexpectedly, “woman” did not exhibit this signature but instead resembled control concepts, such that the independence of the two dimensions did not seem so salient for “woman.”

## 4. Experiment 3

So far, we have shown that in Experiment 1, participants readily made judgments about “good” versus “true” statements for gender concepts just as they did for dual-character concepts. In Experiment 2, we also saw that “man,” although not “woman,” resembled dual-character concepts in participants’ relatively high and equal endorsement of the independence of the concrete and abstract dimensions. We now turn to the next two experiments in Knobe et al. (2013) in which they moved away from simply asking about whether the relevant statements *sounded* natural without any background information. Instead, they examined how participants would judge category membership of specific individuals, depicted in vignettes, who had sufficient traits on the concrete dimension but lacked the core features on the abstract dimension. The main purpose of this measure

was to examine whether violating the abstract norms or lacking the abstract qualities associated with a category could rule someone out from the category despite having sufficient concrete traits. Knobe et al. (2013) found that after reading vignettes about such individuals, only for dual-character concepts would participants endorse both the statement “there is a sense in which X is clearly a... (e.g., artist)” (member statement) and the statement “ultimately...you’d have to say X is not truly a... (artist)” (nonmember statement). For control concepts, Knobe et al. drafted vignettes that had analogous violations of relevant norms (e.g., a cashier who does not care about doing the math right), but after reading the vignettes, participants did not endorse the statement “ultimately...you’d have to say X is not truly a... (cashier).” Following Knobe et al. (2013), we drafted the gender vignettes with violations of well-documented gender norms to establish the abstract dimension of gender concepts and examined whether gender concepts would resemble dual-character concepts in that participants would endorse that the character in the vignette is not truly a man or woman based on the violations. That is, for example, would they agree both that “there is a sense in which the character is clearly a man” and that “ultimately...you’d have to say the character is not truly a man”?

#### 4.1. Method

##### 4.1.1. Participant

We recruited 153 participants on Amazon Mechanical Turk for Experiment 2. The final sample included 147 participants, as six participants did not answer the attention check question correctly ( $M_{\text{age}} = 37.24$ ; age range: 19–77; 59.2% male; 40.8% female).

##### 4.1.2. Materials and procedure

We adopted the same vignettes from Knobe et al. (2013) for dual-character and control concepts and created our own vignettes for gender concepts. All the vignettes described someone or something that had sufficient concrete traits but lacked the abstract values/features associated with a certain category.

Extending this paradigm to gender, we created vignettes based on research examining prescriptive norms for “man” (Vandello et al., 2008) and for “woman” (Eagly & Karau, 2002; Heilman, 2001). Specifically, the “man” vignette was “John is the father of two children. However, John is very emotional, as he cries for many small things in everyday life. John is also a stay-at-home dad who has never contributed anything to the family’s finance and is very submissive to his wife. Moreover, he never sticks up for himself or his family when challenged by others.” The “woman” vignette was “Linda is the mother of two children. However, Linda is very dominant, as she never shows even the slightest hint of weakness and is considered intimidating by many of her colleagues in construction. Linda is the breadwinner of her family and is very bossy with her husband. Moreover, she always sticks up for herself and her family when challenged by others.” It is important to note that in our gender vignettes, the protagonists’ gender identity was directly given by being a mother or father with feminine or masculine pronouns, and the

protagonists were portrayed as transgressing some of the most important abstract gender norms stereotypic to his or her gender identity.

Participants received all of the 32 vignettes in a randomized order, and after reading each vignette, they first rated the concrete member statement and then the “ultimately a nonmember” statement on a 7-point Likert scale ranging from 1 (*false*) to 7 (*true*). In addition to the demographic questions about participants’ gender and age, we also asked about participants’ political orientation using a binary measure on which participants chose to describe themselves as either “more conservative” or “more liberal.”

## 4.2. Results and discussion

### 4.2.1. LMER models

As in previous experiments, we first compared “man” and “woman”: There was a significant interaction between concept and statement type,  $b = -1.18$ ,  $SE = 0.16$ ,  $t(292) = -7.59$ ,  $p < .001$ , such that even though for both “man” and “woman,” member statements were rated much higher than were nonmember statements, the gap between ratings for the two statements was larger for “woman,”  $b = 5.18$ ,  $SE = 0.17$ ,  $t(221.7) = 29.92$ ,  $p < .001$ , than for “man,”  $b = 4.00$ ,  $SE = 0.17$ ,  $t(221.7) = 23.12$ ,  $p < .001$ . We therefore separated “man” and “woman” in subsequent analyses.

We then examined whether the four types of concepts differed significantly in how participants rated the two types of statements for the concepts. The results revealed a significant interaction between concept and statement types,  $F(3, 18.43) = 8.85$ ,  $p < .001$  (see Fig. 3). Planned contrasts first showed that the results in Knobe et al. (2013) successfully replicated: The difference between participants’ ratings of the two types of statements differed significantly for dual-character and control concepts,  $b = 2.62$ ,  $SE = 0.63$ ,  $t(19.24) = 4.14$ ,  $p < .001$ . Specifically, for dual-character concepts, nonmember statements did not differ significantly from member statements,  $b = -0.14$ ,  $SE = 0.46$ ,  $t(21.0) = 0.30$ ,  $p = .77$ , whereas for control concepts, nonmember statements had significantly lower ratings than did member statements  $b = -2.76$ ,  $SE = 0.46$ ,  $t(22.0) = 5.96$ ,  $p < .001$ .

We then conducted planned contrasts to examine how gender concepts compared to other types of concepts. The results showed that both “man” and “woman” were similar to control concepts but significantly different from all other types of concepts in how the two types of statements differed from each other in participants’ endorsement (see Table 1 and Fig. 3).

Given that the gender vignettes reflected perspectives about traditional gender norms and roles, participants’ socio-political orientation might have played a role in their responses (Lindsey, 1990). We therefore examined whether participants’ political orientation (33.6% more conservative; 66.4% more liberal) moderated participants’ endorsement of the two statements for “man,” “woman,” dual-character, and control concepts. The results showed that participants’ political views did not significantly moderate how they endorsed the two statements differently for the four types of concepts,  $F(3, 213.8) = 0.26$ ,  $p = .86$ , nor did their political views significantly moderate their



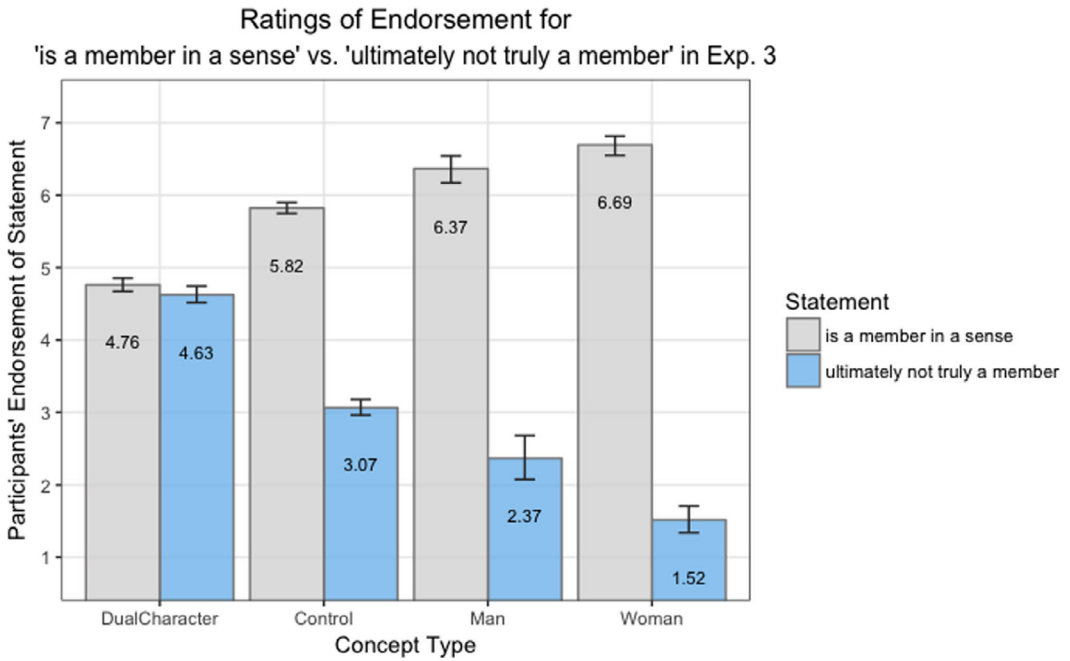


Fig. 3. Participants’ ratings of the “is a member in a sense” versus “not truly a member” statements for dual-character concepts, control concepts, man, and woman in Experiment 3. Error bars represent bootstrapped 95% confidence intervals.

Table 1  
Planned contrasts between dual-character, control, man, and woman in Experiment 3

Comparison	Concept Type × Statement Interaction	Main Effect of Concept Type
Dual-character versus Man	$b = 3.86, SE = 1.46, t = -2.65, p = .02$	$b = 0.33, SE = 0.26, t = 1.26, p = .22$
Control versus Man	$b = -1.24, SE = 1.46, t = -0.85, p = .41$	$b = 0.08, SE = 0.26, t = 0.30, p = .77$
Dual-character versus Woman	$b = 5.04, SE = 1.46, t = 3.45, p = .003$	$b = 0.59, SE = 0.26, t = 2.26, p = .04$
Control versus Woman	$b = -2.42, SE = 1.46, t = -1.66, p = .11$	$b = 0.34, SE = 0.26, t = 1.31, p = .21$

endorsement of the two statements for “man” alone,  $F(1, 288) = 0.198, p = .66$ , or “woman” alone,  $F(1, 288) = 0.203, p = .65$ .

#### 4.2.2. Bayes factors

As in previous experiments, we obtained Bayes factors for a set of regression models representing the hypotheses about how the concept types compare to one another (see Table C in Section 4.1 in Appendix S1). The results, consistent with the findings from

the LMER models, showed that control concepts were the closest to both “man” and “woman,” as the Bayesian models that equated “man” or “woman” with control concepts yielded the largest Bayes factors, which exceeded, by large margins ( $\text{BFs} > 7.94 \times 10^{55}$ ), the Bayes factors from all other models that equated “man” or “woman” with dual-character concepts. Moreover, consistent with the finding regarding the difference between “man” and “woman” in the LMER models, the Bayesian model that equated “man” with dual-character concepts had a larger Bayes factor than the model that equated “woman” with dual-character (by a Bayes factor of  $2.47 \times 10^{44}$ ), suggesting that “man” was still closer than “woman” to dual-character concepts in this case.

### 4.3. Conclusion

In summary, the results from Experiment 3 suggested that when participants based their judgments on vignettes where the individual possessed sufficient traits on the concrete dimension but lacked core features on the abstract dimension, surprisingly neither “man” nor “woman” gender concepts resembled dual-character concepts. We conducted a follow-up study to explore whether a social desirability bias could have been influencing these results. Perhaps participants did not want to explicitly disqualify someone as being truly a man or truly a woman to avoid seeming sexist or narrow-minded. In the follow-up experiment we used the same vignettes but changed the questions to ask participants about how *society*, rather than the participants themselves, would judge the individuals. We found that the results remained the same such that “man” and “woman” resembled control concepts rather than dual-character concepts (see <https://osf.io/f5y2q/> for detailed results).

One possibility for our findings might be the nature of the vignettes we used for gender concepts. Although we chose gender traits that were specified in the modern literature, we wondered if the traditional, stereotypic traits and gender norms that we chose for the abstract dimensions of “man” and “woman” might not be the ones that participants considered central to the abstract dimensions. If so, this might help explain the failure to find the signature of dual-character concepts for gender concepts in Experiments 3. Therefore, we conducted subsequent experiments to probe this possibility.

## 5. Experiment 4

Why did the signature of dual-character concepts disappear for both male and female gender concepts once the judgments were about specific individuals? Before probing further into this question, for the sake of completeness, we replicated the fourth experiment in Knobe et al. (2013) that also used vignettes but asked different forms of questions than those in their Experiment 3 due to the possibility that linguistic hedges such as “ultimately” and “there’s a sense” in the statements in Experiment 3 might have influenced participants’ judgments. They found, however, that removing the hedges did not change participants’ judgments. To provide as complete an assessment as possible, we followed

Knobe et al.'s (2013) approach to determine whether the results in our Experiment 3 would change without these linguistic hedges and found that the pattern did not change (see Section 5 in Appendix S1 for specifics of method and results).

## 6. Experiment 5

Since participants did not think that the transgressions of well-established gender norms used in the previous vignettes would disqualify the protagonists from their gender membership, we speculated that participants might have held different views from traditional gender norms of what it means to be a “true man or woman.” We therefore sought in this experiment to elicit participants’ own views of the circumstances under which they would say someone is not a true man or true woman. Specifically, in this experiment, we asked participants to first rate whether they thought it was natural to say a “true man” and “true woman,” and if they thought it was natural, they were then asked to provide their responses to open-ended questions about the circumstances under which they would say someone is not a true man or a true woman.

### 6.1. Method

#### 6.1.1. Participants

We recruited 411 participants on Amazon Mechanical Turk. After the removal of participants ( $N = 26$ ) who did not provide relevant, comprehensible responses, the current experiment included 385 participants ( $M_{\text{age}}$ : 35.58; age range: 19–71; 48% female; 52% male), 195 of whom were randomly assigned to respond to questions on “true man” and 190 of whom to “true woman.”

#### 6.1.2. Materials and procedure

Similar to Experiment 1, participants first rated two dual-character concepts (artist and scientist) and two control concepts (doorman and cashier), presented in a randomized order, on whether the statement “true X” sounded natural for each concept, using a 7-point Likert scale (1—*sounds weird* to 7—*sounds natural*). At the end of these ratings, participants were randomly assigned to rate either the “true man” or “true woman” statement on the perceived naturalness. If their rating was below 4, they were then prompted to give their open-ended response to why they did not think the statement sounded natural. If their rating was equal to or greater than 4, they were then prompted to give an open-ended response detailing the circumstances under which they would say someone is not a true man or a true woman.

### 6.2. Results and discussion

We first assessed participants’ ratings of the perceived naturalness of the “true man” and “true woman” statements. Replicating our findings in Experiment 1, the results

showed that participants perceived the “true” statements to sound sufficiently natural for both “man” and “woman.” Specifically, one-sample  $t$  tests (two-tailed) showed that the mean ratings for both “man” and “woman” were significantly larger than the midpoint (4) of the 7-point Likert scale (man:  $M = 5.42$  [95% CI: 5.21~5.65],  $t(205) = 12.71$ ,  $p < .001$ ; woman:  $M = 5.09$  [95% CI: 4.85~5.35],  $t(204) = 8.70$ ,  $p < .001$ ). We then calculated the percentage of participants who gave a rating equal to or greater than 4, and the results showed that among those who responded to the “true man” statement, 87% ( $N = 170$ ) gave a rating equal to or higher than 4, and among those who responded to the “true woman” statement, 78% ( $N = 151$ ) gave such a rating. We then coded these participants’ open-ended responses as to the circumstances under which they would say someone is not a true man or true woman.

The coding categories for participants’ open responses were not pre-imposed. Instead, we identified the major themes that emerged from the qualitative data. Specifically, there emerged five major themes that broadly captured all participants’ responses. Specifically, for both “man” and “woman,” the major themes in participants’ responses were, based on overall frequency, (a) moral transgressions (e.g., being abusive, cheating on partner), (b) unspecified gender nonconformity (e.g., men being feminine, women being masculine), (c) identity-relevant judgments (e.g., being transgender, non-adults, robots), (d) stereotypic gender norm transgressions (e.g., men being submissive, women being aggressive), and (e) others (e.g., not able to think of a specific circumstance despite high rating of the statement). The themes were not mutually exclusive such that the same response might contain multiple themes. For the circumstances under which participants would say someone is not a “true man,” moral norm transgressions were surprisingly the most prevalent theme in participants’ responses, accounting for 47% of the total responses given, followed by “identity”-based judgments (25%), stereotypical gender norm transgressions (14%), unspecified gender nonconformity (14%), and others (7%). In contrast, for “true woman,” the most prevalent theme was “identity”-based judgments, accounting for 48% of the total responses given, followed by moral norm transgressions (24%), unspecified gender nonconformity (21%), stereotypical gender norm transgressions (13%), and others (8%). The proportions of the themes differed significantly for “man” and “woman,”  $\chi^2(4, N = 352) = 26.06$ ,  $p < .001$ .

The results from the open responses unexpectedly suggested that participants in general considered moral transgressions, rather than stereotypic gender norm transgressions, to be definitional of being a “true man.” For “true woman,” as well, stereotypic gender norm transgressions were not a frequently mentioned defining feature. Aside from identity-based judgments and global, unspecified gender nonconformity, only moral transgressions showed up with any frequency.

If vignettes emphasized moral transgressions instead of stereotypical traits, would people now judge that a man or woman committing the transgression was not “truly a man or woman?” We examined this question in the following experiments.

## 7. Experiment 6a

In Experiment 6a, we sought to support the finding from participants' open responses in Experiment 5 by examining how moral transgressions might compare to gender norm violations in participants' judgments of how serious each type of violation was in terms of their image of the ideal man or woman. In this way, we continued our search for the potential norms and traits that might define the abstract dimension of gender concepts. Moreover, since previous research has shown that judgments about category ideals are different from judgments about category membership per se (Barsalou, 1985), we sought to examine whether the same transgressions that would likely lead to negative judgments about whether someone was an "ideal man or woman" would or would not align with judgments about being "truly a man or woman"—with only the latter being a marker of a dual-character concept.

### 7.1. Method

#### 7.1.1. Participants

We recruited 204 participants ( $M_{\text{age}}$ : 34.87; age range: 18–69; 43% female; 56% male; 1% not reported) on Amazon Mechanical Turk, 101 of whom received items for "man" and 103 of whom received items for "woman."

#### 7.1.2. Materials and procedure

Participants in both "man" and "woman" conditions were presented with a randomized series of 23 statements about different individuals, each differing in content (*moral*: e.g., committing abuse, being dishonest; *gender stereotypic*: e.g., a woman being dominant, a man being submissive; and *positive-valence manipulation check*: e.g., curious, loves traveling). The moral traits and acts were selected based on the prevalent themes in participants' morality-related responses in Experiment 5, with "abusing children and partner" and "cheating on partner" being the most prevalent ones for both "man" and "woman." Gender-stereotypic traits and acts were selected based on those portrayed in the gender vignettes we created in Experiment 3. There were two versions of moral and gender-stereotypic statements that differed in specificity (general trait adjectives vs. specified actions). Specifically, half of the participants were randomly assigned to rate moral and gender-stereotypic traits described in general trait-adjective statements (e.g., John is abusive) and the other half were randomly assigned to rate the moral and gender-stereotypic traits described in specified action statements (e.g., John has been abusing his children). All participants received the same manipulation check statements. The goal was to assess whether the specificity of how a moral or gender-stereotypic trait was portrayed affected participants' judgments. After reading each statement, participants rated, on a 7-point Likert scale ranging from 1 (*not a violation at all*) to 7 (*a serious violation*), the extent to which they considered the trait or act described in the statement to be a serious violation of their ideal images of a man or woman.

## 7.2. Results and discussion

We first examined whether the specificity of how transgression of a certain moral or gender-stereotypic trait was portrayed in the statements affected participants' judgments. The results showed that there was no significant difference between participants' ratings of general adjective (trait) statements and specified actions statements, for either "man,"  $b = -0.58$ ,  $SE = 0.34$ ,  $t(44.90) = -1.68$ ,  $p = .10$ , or "woman,"  $b = -0.45$ ,  $SE = 0.38$ ,  $t(63.17) = -1.20$ ,  $p = .23$ . The results also showed that transgressions of moral and gender-stereotypic norms were rated as significantly more serious violations than the manipulation check items, both when portrayed in general adjective statements ["man":  $b = 2.44$ ,  $SE = 0.38$ ,  $t(49.12) = 6.47$ ,  $p < .001$ ; "woman":  $b = 1.96$ ,  $SE = 0.39$ ,  $t(55.40) = 5.01$ ,  $p < .001$ ] and when portrayed in concrete act statements ["man":  $b = 3.01$ ,  $SE = 0.38$ ,  $t(50.36) = 7.94$ ,  $p < .001$ ; "woman":  $b = 2.42$ ,  $SE = 0.37$ ,  $t(52.14) = 6.51$ ,  $p < .001$ ]. Since the specificity of the statements that described transgressions of moral and gender-stereotypic norms did not influence participants' judgments, we therefore in the subsequent analyses will focus on the eight most highly rated statements (see Fig. 4), regardless of specificity, for moral and gender-stereotypic items, respectively. If the two types of statements of the same transgression both appeared among the most highly rated statements, we will focus only on the more highly rated of the two—for example, if both "John

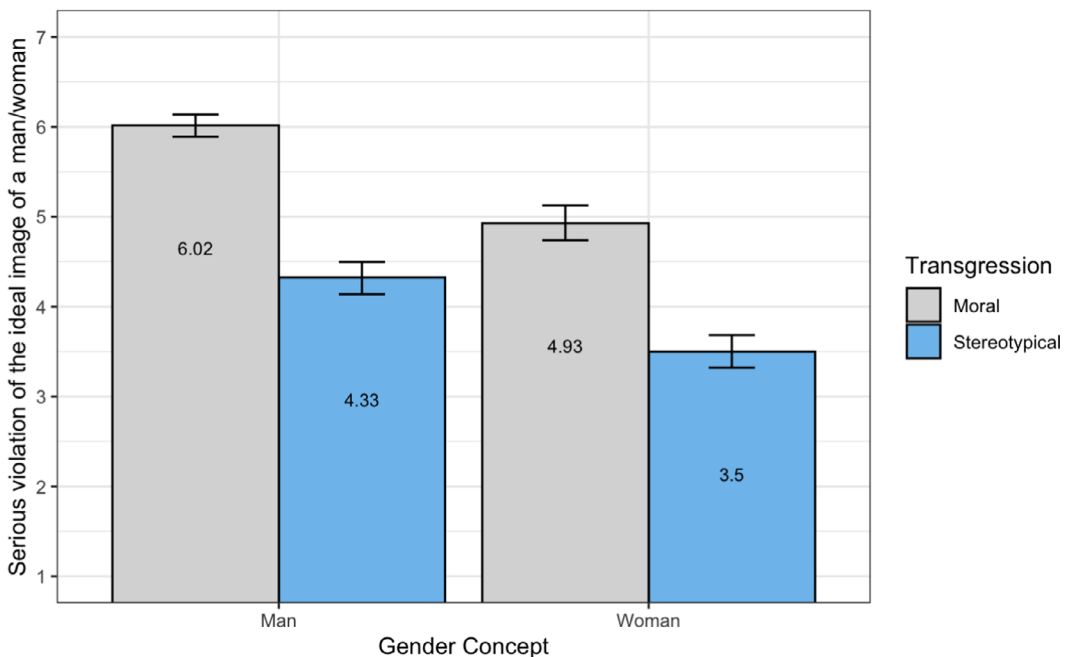


Fig. 4. Participants' ratings of how serious each specific transgression was as a violation of their ideal image of a man versus a woman in Experiment 6a. Error bars represent bootstrapped 95% confidence intervals.

is abusive” and “John has been abusing his children” were among the most highly rated statements, we will only focus on the more highly rated one.

As shown in Fig. 5, comparing moral against stereotypic gender transgressions, we found that when judging both “man” and “woman,” participants considered moral transgressions to be significantly more serious violations of the ideal images of a man and woman than they considered stereotypic gender transgressions [“man”:  $b = 1.57$ ,  $SE = 0.36$ ,  $t(17.30) = 4.40$ ,  $p < .001$ ; “woman”:  $b = 1.40$ ,  $SE = 0.39$ ,  $t(16.19) = 3.56$ ,  $p < .003$ ]. In addition, for both “man” and “woman,” moral transgressions on aggregate were rated significantly higher than the midpoint (4) on the 7-point Likert scale, suggesting high level of perceived seriousness of these transgressions as violations of participants’ ideal images of a man or woman [“man”:  $M = 5.94$  [95% CI: 5.82~6.07],  $t(397) = 30.20$ ,  $p < .001$ ; “woman”:  $M = 4.93$  [95% CI: 4.75~5.12],  $t(403) = 9.80$ ,  $p < .001$ ]. Moreover, although the general pattern that moral transgressions were considered more serious than stereotypic gender transgressions did not differ for “man” and “woman” (i.e., no interaction),  $F(1, 38.37) = 0.02$ ,  $p = .89$ , both moral transgressions and stereotypic gender transgressions alone were considered more serious violations of participant’s ideal images of a man than their ideal images of a woman [moral:  $b = 5.39$ ,  $SE = 0.24$ ,  $t(29.3) = 22.54$ ,  $p < .001$ ; stereotypic:  $b = 3.91$ ,  $SE = 0.21$ ,  $t(34.7) = 18.92$ ,  $p < .001$ ].

### 7.3. Conclusion

The findings in the current experiment supported our conjecture that participants might not have considered the traditional, stereotypic gender norms used in previous experiments to possess sufficient normative force to disqualify someone from their gender group. Indeed, participants, for both “man” and “woman,” considered moral transgressions to be much more serious violations than gender norm transgressions of their ideal images of a man or woman. Would participants in fact go so far as to disqualify a person of his or her membership as truly a man or truly a woman on the abstract dimension based on these moral transgressions? We conducted Experiment 6b to examine this interesting possibility.

## 8. Experiment 6b

In Experiment 6a, we found that participants indeed perceived moral transgressions to be significantly more serious than stereotypic gender transgressions as violations of their *ideal image* of both a man and a woman. However, would participants actually disqualify someone as being a “*truly* a man” or “*truly* a woman” based on these moral transgressions? To answer this question, we conducted the next experiment.

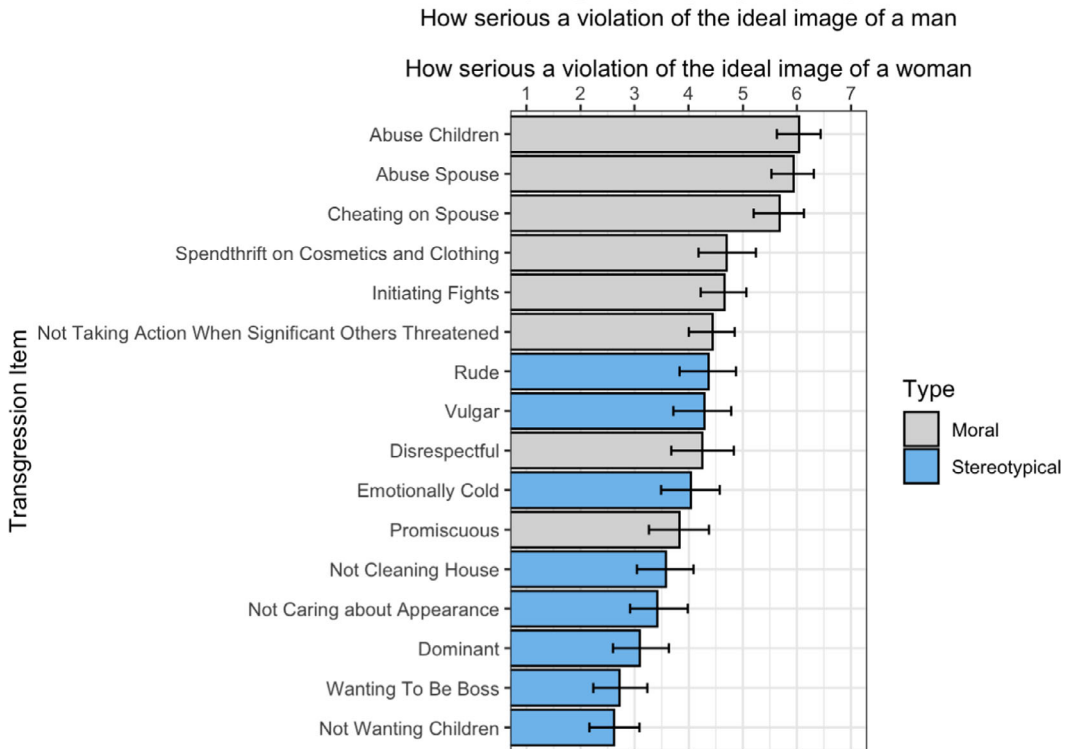
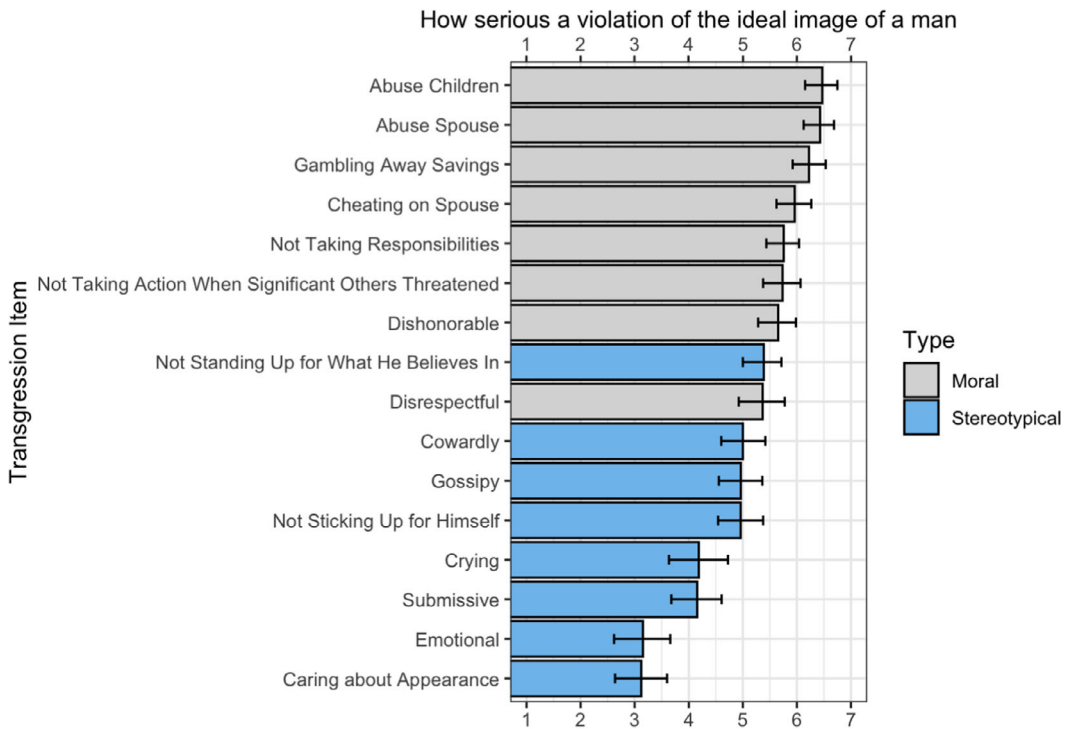




Fig. 5. Participants' aggregate ratings of how serious transgressions of moral versus gender-stereotypic norms were as violations of their ideal image of a man versus a woman in Experiment 6a. Error bars represent bootstrapped 95% confidence intervals.

---

## 8.1. Method

### 8.1.1. Participants

We recruited 303 participants on Amazon Mechanical Turk ( $M_{\text{age}} = 33.19$ ; age range: 18–74; 40% female; 59% male; 1% nonbinary), 151 of whom received items for “man” and 152 of whom received items for “woman.”

### 8.1.2. Materials and procedures

We used the same items for moral and stereotypic gender transgressions as in Experiment 6a for the current experiment. This time, however, we asked participants to make explicit judgments, on a 7-point Likert scale ranging from 1 (*false*) to 7 (*true*), about the statement, “in this sense, X is not truly a man or woman,” after reading about the character violating either a specific moral norm (e.g., abusing children) or a gender-stereotypic norm (e.g., being emotional for “man” or bossy for “woman”).

## 8.2. Results and discussion

Since participants judged whether someone is not “truly a man” or “truly a woman” based on the same transgressions adopted from Experiment 6a, their judgments about being a “truly a man or woman” in the current study might be influenced by how serious the transgressions were considered to be, as assessed in Experiment 6a, in terms of violating the ideal images of a man or woman. For this reason, in the current experiment, we took the average score of each transgression item in Experiment 6a as a covariate (both as an additive fixed effect and a by-item random effect) in the subsequent LMER analyses.

Comparing moral against stereotypic gender transgressions, we first found that the degree to which each transgression item was considered to be a serious violation of the ideal images of a man or woman as assessed in Experiment 6a indeed significantly affected participants' ratings in the current experiment, as the model that included the ratings from Experiment 6a as a covariate was significantly better than the model without the ratings,  $\chi^2(3) = 45.65$ ,  $p < .001$ . Nonetheless, even with the covariate's significant influence accounted for, other variables of interest in the current experiment remained independently significant (see Fig. 6). Specifically, participants' ratings for moral versus stereotypic gender transgressions differed greatly for “man” as compared to “woman,”  $b = 1.19$ ,  $SE = 0.15$ ,  $t(82.69) = 7.68$ ,  $p < .001$ . For “man,” moral transgressions were considered significantly more disqualifying than stereotypic gender transgressions of being “truly a man,”  $b = 1.17$ ,  $SE = 0.14$ ,  $t(50.8) = 8.63$ ,  $p < .001$ , whereas for “woman,” moral transgressions—although they were considered to be serious for women

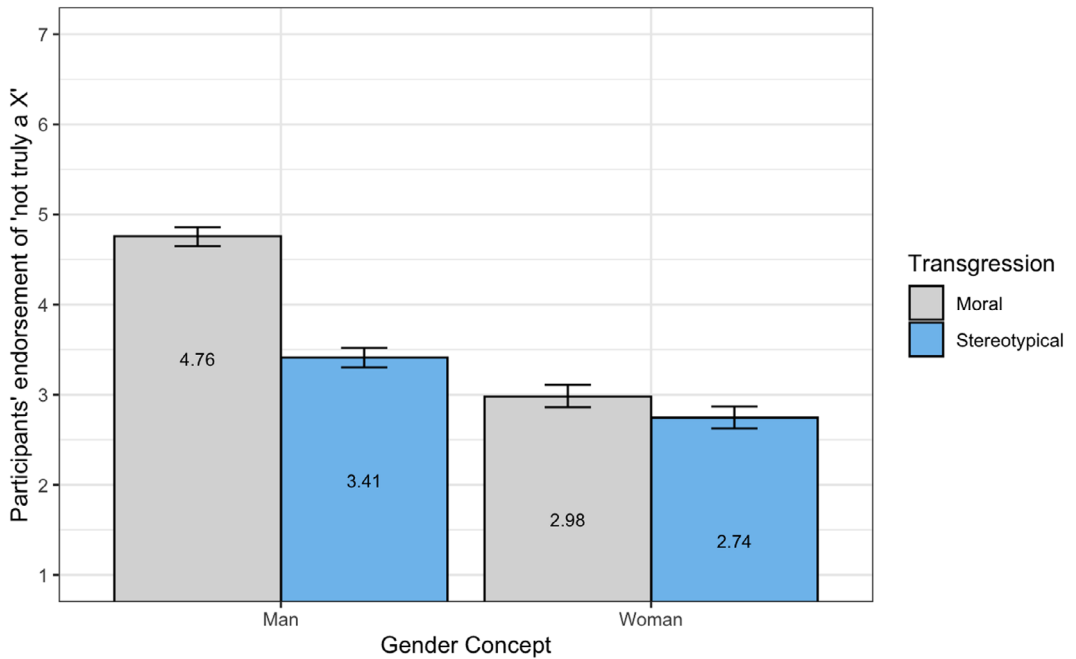
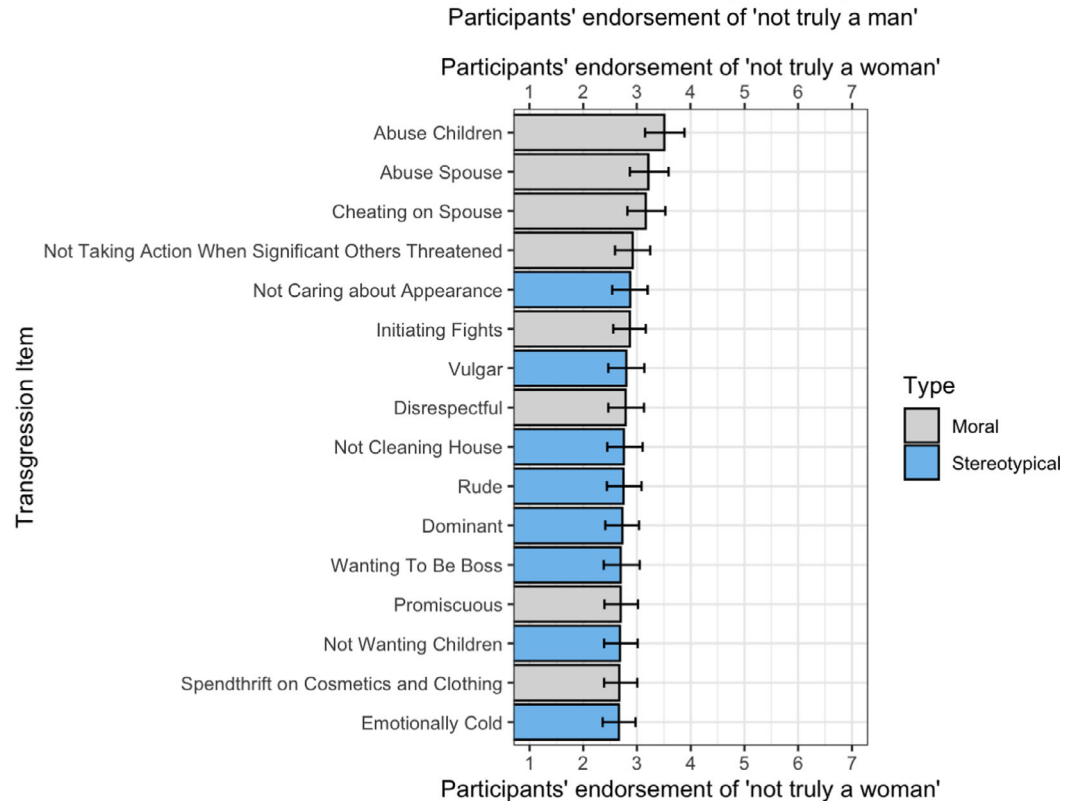
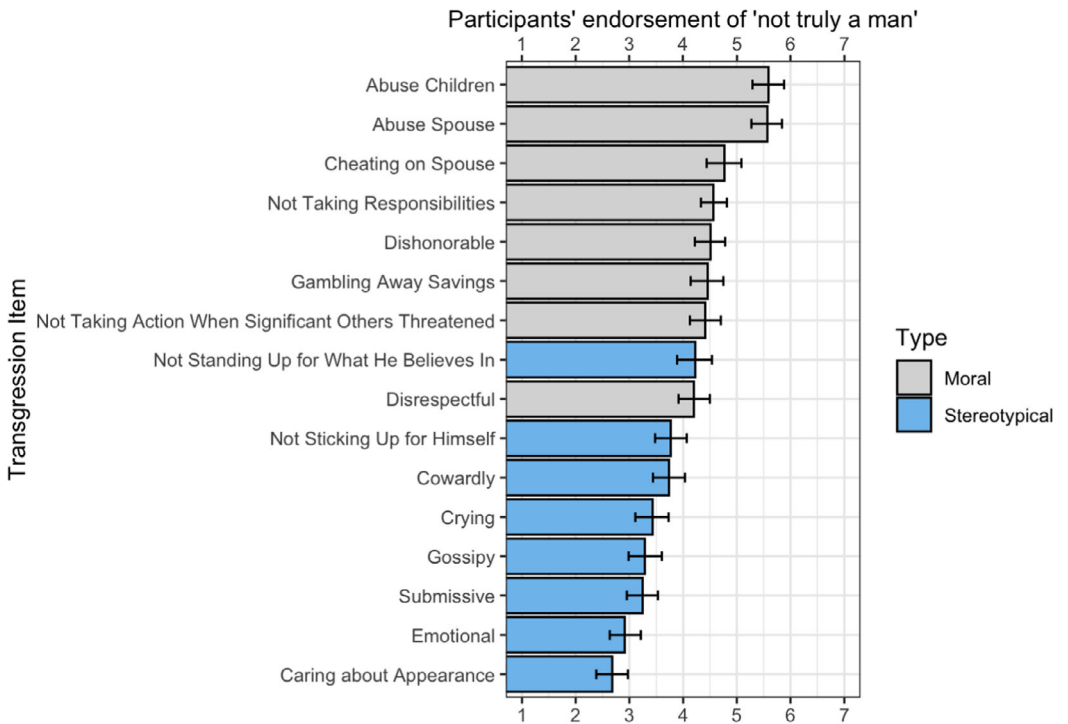


Fig. 6. Participants' endorsement of "X is not truly a man or woman" averaged across moral versus stereotypical gender norm violations in Experiment 6b. Error bars represent bootstrapped 95% confidence intervals.

in Experiment 6a—did not differ from stereotypic gender transgressions in the extent to which they were considered disqualifying of being "truly a woman,"  $b = -0.02$ ,  $SE = 0.11$ ,  $t(21.9) = -0.14$ ,  $p = .89$ .

In fact, for "woman," no single transgression item was rated above midpoint (4) on the 7-point Likert scale assessing how disqualifying of being "truly a woman" each specific transgression is (see Fig. 7). Together, the ratings of moral transgressions for disqualifying "woman" were significantly below the midpoint (see Fig. 6),  $M = 2.98$  [95% CI: 2.86~3.10],  $t(1215) = -16.82$ ,  $p < .001$ . In contrast, for "man," every moral transgression was rated above the midpoint (see Fig. 7), and the moral transgressions together were rated significantly higher than the midpoint for disqualifying a man from being "truly a man" (see Fig. 6),  $M = 4.76$  [95% CI: 4.65~4.87],  $t(1207) = 13.80$ ,  $p < .001$ .

Fig. 7. Participants' endorsement of "X is not truly a man or woman" based on each specific norm transgression in Experiment 6b. Error bars represent bootstrapped 95% confidence intervals.



### 8.3. Conclusion

The results of the current experiment showed that participants indeed thought that moral transgressions, rather than stereotypic gender transgressions, could disqualify a man from being *truly* a man. In contrast, participants did not think either moral or stereotypic gender transgressions would disqualify a woman from being *truly* a woman, despite the high ratings for moral transgressions in the previous experiment as serious violations of participants' *ideal image* of a woman. Therefore, replacing gender-stereotypic norms with moral norms as the central criterion of the abstract dimension restored the dual-character signature for "man" but not for "woman." In other words, violating certain moral values and transgressing certain moral norms could indeed disqualify a man as being "truly a man" but did not disqualify a woman from being "truly a woman."

One might think that perhaps participants interpreted "man," but not "woman," as referring to the generic "human," as in "no man is an island" and "all men are created equal." Since moral traits are human characteristics, the generic meaning of "man" as "human" might explain why moral transgressions disqualified being "truly a man" but not "truly a woman." However, while there are certainly contexts where "man" can be understood as referring to the generic "human," we do not think this is a plausible explanation for our results.

Specifically, there were multiple salient cues in Experiments 6a and 6b that clearly indicated that "man" referred to the male gender rather than a generic for "human." First, participants read and rated multiple statements that each depicted a specific person with a male name and male pronouns, such as "Henry cries for trivial things in everyday life. In this sense, he is not truly a man" and "Jason has been abusing his children. In this sense, he is not truly a man." It should therefore be clear to participants that "man" here refers to the male gender. Second, since participants saw, in a randomized order, equal numbers of statements indicating moral transgressions and statements reflecting well-known masculine gender norms, it should be salient to participants that the purpose of the experiment is to judge the impact that these violations would have on the protagonists' gender status as "truly a man." Additionally, we picked the moral items in Experiments 6a and 6b from participants' open responses in Experiment 5. When asked to provide open responses to the question "under what circumstances would you say someone is not truly a man (or woman)?," participants not only frequently mentioned moral transgressions for "man" but also significantly for "woman," where the issue of "human" did not arise. Therefore, we do not think that participants interpreted "man" as meaning the generic "human" in Experiments 6a and 6b.

## 9. Discussion

### 9.1. Overview

Building on groundbreaking research by Knobe et al. (2013), we examined the extent to which the dual-character framework can provide insights into cornerstone social

concepts, in particular “man” and “woman.” Specifically, we examined whether “man” and “woman” could function as dual-character concepts, such that gender identity (i.e., the concrete dimension) and gendered values and norms (i.e., the abstract dimension) could each lead to distinct category membership judgments.

In contrast to the results of Knobe et al., who found consistent patterns for dual-character concepts across measures, we found somewhat different patterns for gender concepts such that (a) while both “man” and “woman” resembled dual-character concepts on metalinguistic judgments, their dual-character signature disappeared on judgments about individuals transgressing well-established traditional gender norms, and (b) “man,” but not “woman,” regained its dual-character nature when moral norms replaced gender norms as the content of the abstract dimension being transgressed. An interesting possibility is that these inconsistencies and differences are mirroring historical shifts in conceptualizations of gender concepts. Later, we also speculate on how social changes may be important not only for dual-character concepts, which are partly constituted by social norms and values, but also for representations of other types of concepts, such as artifacts and ontological categories.

## *9.2. Implications for the dual-character concept framework*

### *9.2.1. Metalinguistic judgments versus judgments about specific qualities*

As just mentioned, unlike the consistent patterns for dual-character concepts across the measures in Knobe et al.’s (2013) original studies, our examination of gender concepts revealed that the metalinguistic judgments, such as the perceived naturalness of statements containing the phrase “true man” or “true woman,” could differ from judgments based on the embodiment of abstract norms and values. We do not view these inconsistencies across measures as flaws of the original paradigm but rather as revealing meaningful additional information about dual-character concepts and their evolution. In particular, the contradictory results for “woman” in metalinguistic and vignette-based tasks might reflect a difference between linguistic intuitions about how a category used to be represented and assessment of how a category is now mentally represented. Specifically, metalinguistic judgments, such as whether it sounds natural to say, “a true woman,” might rest on using the phrase “true woman” as a placeholder for the abstract dimension but in this case the content of the abstract dimension was not specified (Malt, 1990). In contrast, for both “man” and “woman,” once the abstract dimension was depicted in vignettes with detailed portrayals of gender norm transgressions, participants did not disqualify the men/women of being “truly a man or woman.” When the depictions shifted instead to moral norms in the vignettes, “man” regained its dual-character status, but “woman” did not, despite the fact that moral transgressions, such as neglecting your children, were considered egregious for women as well. The judgments based on specific qualities therefore revealed important information about the possible evolution of how gender concepts are represented.

Metalinguistic judgments and vignette-based judgments have been separately used as major methods in empirical studies in the field of psycholinguistics and in the study of concepts (Keil, 1989; Malt, 1990). However, the use of both methods for examining gender concepts in our studies suggested that the two methods may in fact assess different conceptual representations. Metalinguistic judgments, for example, that “true woman” sounds felicitous, are devoid of specific content and may reflect the strong sense that there is an abstract dimension to a concept, but when the content is fleshed out in a vignette, judgments about whether the specific attributes disqualify someone from being “truly a woman” may reflect currently held beliefs about whether those qualities are considered central to the abstract dimension. Therefore, using the two methods in tandem might uncover important information about concepts that have undergone significant historical change and thus have different past versus present representations.

### 9.2.2. *Historical change and representations of dual-character concepts*

Why were *gender* norm transgressions no longer considered disqualifying for being “truly a man/woman”? One possibility, as previously suggested, is that to the extent that stereotypic gender norms still exist, they may no longer serve as the criteria for the abstract dimension for gender concepts in more recent times (see Marcus, Page, Calder, & Foley, 2014 for a review of literature on change in gender norms). Many members of the younger generation seem prone to resist traditional gender norms; for example, men are increasingly engaged in traditionally feminine activities such as child-rearing and housework (Vandello & Bosson, 2013). Similarly, women’s participation in traditionally male-dominated fields has increased significantly: The proportion of women in the workforce of all STEM fields increased from 7% in 1970 to 26% in 2018, and notably, in math occupations alone, from 15% in 1970 to 48% in 2018 (U.S. Bureau of Labor, 2019).

Changes and advances in social trends and policies have been documented to affect people’s normative perception. For example, a recent longitudinal study on people’s perceptions of sexual orientation showed that the legalization of same-sex marriage in 2015 dramatically shifted people’s perceptions of social norms—and acceptance of homosexuality was now perceived to be the social norm of the current American society (Tankard & Paluck, 2017). Similarly, there has been a gradual decline in both implicit and explicit bias against homosexuality, which became significantly sharper after the federal legalization of same-sex marriage in 2015 (Ofosu, Chambers, Chen, & Hehman, 2019). Therefore, given how changes in social trends and policies influenced people’s normative perceptions of homosexuality, it is possible that the comparable social changes in, and challenges to, gender stereotypes and roles (e.g., as previously reviewed, people’s increasing engagement in activities and occupations traditionally associated with the other gender) might have also changed lay conceptualizations of what is central to gender membership on the abstract dimension.

Another related but yet unexamined possibility is that people may have gradually construed the nature of gender norms as only conventional (Ayala, 2010; Levy, Taylor, & Gelman, 1995; Pinker, 2008), and now hold gender membership to a higher ground, one

that might resemble the moral high ground (see Knobe, 2010, for how moral concerns may affect seemingly nonmoral judgments). Of course, people may very well continue to police their own behavior based on conformity to traditional gender norms, but the representations of gender categories as concepts may have changed. Turning back to our findings about “man,” there has been evidence showing the emerging importance of morality in the conceptualization of “man”—a recent public opinion poll conducted in the United States by the Pew Research Center showed that honesty/morality, rather than conventional gender norms, was the most mentioned category in people’s responses to what they think are the traits and characteristics that Americans nowadays value most in men (Parker, Horowitz, & Stepler, 2017).

Interestingly, there was a puzzling difference in our findings on moral norms’ importance for the abstract dimension of gender concepts—participants perceived moral transgressions to be much more disqualifying of being “truly a man” than “truly a woman.” Why did people harshly judge a woman who abused her children but, unlike her male counterpart, did not consider her abusiveness to disqualify her as truly a woman? It becomes even more interesting when we consider the fact that criminality, aggression, and predatory behavior are more associated with men than women (Bailey, LaFrance, & Dovidio, 2018).

It is possible that the *descriptive* stereotype of men being more violent and aggressive is not reflected in what is expected of being “truly a man.” The injunctive norms on the abstract dimension for “man” and “woman,” therefore, can diverge from the descriptive stereotypes people hold. Another possibility is that moral transgressions were more disqualifying of being “truly a man” than a “truly a woman” because of the difference in the precariousness of manhood and womanhood (the ease of being disqualified as “truly a man” or “truly a woman”). A review of previous research on gender conformity concluded that manhood is more precarious than womanhood: Manhood is more frequently questioned and scrutinized (Vandello & Bosson, 2013). Thus, an interesting question is whether the greater precariousness of manhood in the context of gender norms has transferred to morality-based judgments about gender membership, making moral transgressions more disqualifying of being “truly a man” than “truly a woman.”

It is also possible that the abstract dimension for gender concepts has come to function like a placeholder, similar to the idea that the “essence” may be a placeholder in psychological essentialism (Medin & Ortony, 1989). Specifically, in the essentialism literature, it has been established that people often have the belief that some categories have an underlying “essence” that determines category membership but much of the time people do not know what exactly the “essence” is, rendering the “essence” a placeholder whose content is often not specified. In the case of gender concepts, scholars in the humanities such as Judith Butler (Ferber, 2020) have argued that gender categories are historical categories that undergo transformations over time, that we do not know all the meanings that gender concepts entail, and that we should always be open to their new social meanings. Therefore, while the placeholder might have undergone change and been replaced with moral norms for “man,” it may still be devoid of content for “woman,” thus explaining why it was considered natural to say “truly a woman” on metalinguistic judgments but no

specific qualities seemed to disqualify an actual woman of being “truly a woman.” Yet, there might still exist other unexamined domains where transgressions of abstract norms would also disqualify a woman from being “truly a woman.” Therefore, the extent to which “woman” is still a dual-character concept remains an intriguing question.

Our examination of gender concepts thus suggested that changes in representations of dual-character concepts may be manifested not only in what is considered characteristic of the abstract dimension, as in the case of “truly a man,” but also in a concept’s shifting away from being a dual-character concept as the abstract dimension loses its importance in category membership judgments or remains a void placeholder, as it might have with “woman.” As another example, even for dual-character concepts like “artist,” the criteria for category membership on the two dimensions may be evolving. For instance, graffiti used to be considered vandalism, but it is now increasingly perceived to be a form of contemporary art that embodies unique artistic creativity and postindustrial cultural values (McAuliffe, 2012). The extent to which a concept functions as a dual-character concept and the exact content that makes up the abstract dimension, and perhaps even the concrete dimension, of a dual-character concept may therefore not be fixed but may be subject to change over time, particularly for social categories. Examining the changes in mental representations of categories can thus yield important insights about social categories that are inextricably tied to social trends and change.

The dual-character concept framework may also help demonstrate how some of the most profound individual differences, such as different ideologies of social issues, can shape our mental representations of categories. Nationality concepts might be particularly interesting, as concepts such as “American” can refer to either the concrete citizenship or particular cultural values. For instance, “American” can be thought of as having a legal identity as a U.S. citizen, and undocumented immigrants might not qualify as “American” in this sense, but for some people, “American” may be thought of as referring to the endorsement and practice of American shared values, and someone growing up in the United States without documentation might very well qualify as “American” in this sense. Whether “American” is a dual-character concept, therefore, may be contingent on the ideological views of a person. For example, some people might not treat “American” as a dual-character concept, relying only on whether the individual has legal documentation as a U.S. citizen in assessing category membership for “American.” As such, the dual-character concept framework may be able to not only reflect historical changes in conceptual representations but also capture how different ideologies or cultural beliefs can lead to very different conceptual representations of the same category—this is an exciting untouched area worthy of future empirical investigation.

### *9.2.3. Relevance to change in conceptual representations in general*

Such implications are not limited to dual-character concepts but may be true for conceptual representations in general. Other types of concepts might also be more dynamic than previously assumed. For example, our speculation on the historical change of dual-character concepts resonates with research on artifact concepts that challenged the common assumption about category stability. Specifically, Malt and Paquet (2013) showed



that what could be considered “real” instances of artifacts, such as phones and keys, has changed greatly across generations. They found, for example, that compared to older adults, college students were much more likely to categorize mobile phones as real phones and contactless keys as real keys. Similarly, White, Storms, Malt, and Verheyen (2018) demonstrated that when categorizing household containers such as bottles, younger adults regarded features such as “made of plastic rather than glass” as important for “bottle,” but these features did not play a significant role in older adults’ judgments about “bottle.” Such generational differences might be a product of the “plastic age” that started from the 1960s, after which younger generations’ experience with everyday objects was dominated by plastic (rather than, e.g., glass) materials (White et al., 2018).

Likewise, there might also be ongoing changes in what we consider to be ontological concepts, such as sentient and agentic beings, as a result of rapid advancement in robotics and artificial intelligence (Brink, Gray, & Wellman, 2019; Weisman, Dweck, & Markman, 2017). What used to be perceived as pure artifacts, such as robots, might be now increasingly recognized as agentic beings and in the future perhaps even sentient beings (Bigman, Waytz, Alterovitz, & Gray, 2019). Even the boundary between biological beings and artifacts may be becoming blurred by new materials, such as the self-healing “living” concrete that contains bacteria that become active to fill the cracks when they appear and expose the bacteria to air (Wiktor & Jonkers, 2011). These changes in conceptualizations of specific categories might in turn lead us to re-examine the assumptions about the overarching concept types, such as what it means to be an artifact and what it means to be a sentient or biological being.

In conclusion, since the abstract dimension of values and norms is what makes dual-character concepts distinct from most other concepts, understanding changes and transitions in social trends and perceptions can contribute to our understanding of how dual-character concepts come into being and evolve. The effects of the changes are not only limited to the nature of the abstract dimension of dual-character concepts. Rather, the changes can also move a concept in and out of the dual-character concept family, such that what used to be a dual-character concept might no longer belong in the family and vice versa. Moreover, the overarching concept types, such as artifacts and agentic beings, are not fixed either. Like dual-character concepts, a particular concept (or its definitional or characteristic features) might move in and out of a general concept type in accordance with changes in social values and norms or scientific or technical advances. Documenting changes in social trends and norms and tracking their impact on our conceptual representations can therefore be a fruitful avenue of future research, one that may yield important insights into the nature and development of concepts.

## **Acknowledgments**

We thank Joshua Knobe for sharing experimental materials and helpful discussion, and Michael C. Frank for consultation in statistical analyses. This research was partly

supported by a Markowski-Leach Scholarship from Horizons Foundation and a Norman H. Anderson Research Grant from Stanford University to Cai Guo.

## Open Research badges



This article has earned Open Data and Open Materials badges. Data and materials are available at <https://osf.io/f5y2q/>.

## References

- Ayala, F. J. (2010). The difference of being human: Morality. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(Supplement\_2), 9015–9022. <https://doi.org/10.1073/pnas.0914616107>
- Bailey, A. H., LaFrance, M., & Dovidio, J. F. (2018). Is man the measure of all things? A social cognitive account of androcentrism. *Personality and Social Psychology Review*, *23*(4), 307–331. <https://doi.org/10.1177/1088868318782848>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 629–654. <https://doi.org/10.1037/0278-7393.11.1-4.629>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *ArXiv:1506.04967 [Stat]*. Retrieved from <http://arxiv.org/abs/1506.04967>. Accessed February 1, 2018.
- Bayman, E. O., Chaloner, K., & Cowles, M. K. (2010). Detecting qualitative interaction: A Bayesian approach. *Statistics in Medicine*, *29*(4), 455–463. <https://doi.org/10.1002/sim.3787>
- Bigman, Y. E., Waytz, A., Alterovitz, R., & Gray, K. (2019). Holding robots responsible: The elements of machine morality. *Trends in Cognitive Sciences*, *23*(5), 365–368. <https://doi.org/10.1016/j.tics.2019.02.008>
- Brink, K. A., Gray, K., & Wellman, H. M. (2019). Creepiness creeps in: Uncanny valley feelings are acquired in childhood. *Child Development*, *90*(4), 1202–1214. <https://doi.org/10.1111/cdev.12999>
- Eager, C., & Roy, J. (2017). Mixed effects models are sometimes terrible. *ArXiv:1701.04858 [Stat]*. Retrieved from <http://arxiv.org/abs/1701.04858>. Accessed February 1, 2018.
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, *109*(3), 573–598. <https://doi.org/10.1037/0033-295X.109.3.573>
- Eidson, R. C., & Coley, J. D. (2014). Not so fast: Reassessing gender essentialism in young adults. *Journal of Cognition and Development*, *15*(2), 382–392. <https://doi.org/10.1080/15248372.2013.763810>
- Ferber, A. (2020). Judith Butler on the culture wars, JK Rowling and living in “anti-intellectual times”. *New Statesman*. Retrieved from <https://www.newstatesman.com/international/2020/09/judith-butler-culture-wars-jk-rowling-and-living-anti-intellectual-times>. Accessed October 5, 2020.
- Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don’t have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189–211.
- Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York, NY: Oxford University Press.
- Gilmore, D. D. (1990). *Manhood in the making: Cultural concepts of masculinity*. New Haven, CT: Yale University Press.

- Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues*, 57(4), 657–674. <https://doi.org/10.1111/0022-4537.00234>
- Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & van Anders, S. M. (2019). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist*, 74, 171–193. <https://doi.org/10.1037/amp0000307>
- Keil, F. C. (1989). *Concepts, kinds, and conceptual development*. Cambridge, MA: MIT Press.
- Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences*, 33(4), 315–329. <https://doi.org/10.1017/S0140525X10000907>
- Knobe, J., Prasada, S., & Newman, G. E. (2013). Dual character concepts and the normative dimension of conceptual representation. *Cognition*, 127(2), 242–257. <https://doi.org/10.1016/j.cognition.2013.01.005>
- Kohlberg, L. A. (1966). Cognitive-developmental analysis of children's sex-role concepts and attitudes. In E. E. Maccoby (Ed.), *The development of sex differences* (pp. 82–173). Stanford, CA: Stanford University Press.
- Kruschke, J. K. (2015). *Doing Bayesian data analysis*. London, UK: Academic Press. <https://doi.org/10.1016/C2012-0-00477-2>
- Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, 2(4), 458–508.
- Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago, IL: University of Chicago Press.
- Leslie, S. J. (2015). “Hillary Clinton is the only man in the Obama administration”: Dual character concepts, generics, and gender. *Analytic Philosophy*, 56(2), 111–141. <https://doi.org/10.1111/phib.12063>
- Levy, G. D., Taylor, M. G., & Gelman, S. A. (1995). Traditional and evaluative aspects of flexibility in gender roles, social conventions, moral rules, and physical laws. *Child Development*, 66(2), 515–531. <https://doi.org/10.1111/j.1467-8624.1995.tb00887.x>
- Lindsey, L. L. (1990). *Gender roles: A sociological perspective*. New York: Routledge.
- Maccoby, E. E. (1998). In *The two sexes: Growing up apart, coming together*. Cambridge, MA: Belknap Press of Harvard University Press.
- Malt, B. C. (1990). Features and beliefs in the mental representation of categories. *Journal of Memory and Language*, 29(3), 289–315. [https://doi.org/10.1016/0749-596X\(90\)90002-H](https://doi.org/10.1016/0749-596X(90)90002-H)
- Malt, B. C. (1994). Water is not H<sub>2</sub>O. *Cognitive Psychology*, 27(1), 41–70. <https://doi.org/10.1006/cogp.1994.1011>
- Malt, B. C., & Paquet, M. R. (2013). The real deal: What judgments of really reveal about how people think about artifacts. *Memory & Cognition*, 41(3), 354–364. <https://doi.org/10.3758/s13421-012-0270-9>
- Marcus, R., Page, E., Calder, R., & Foley, C. (2014). *Drivers of change in gender norms: An annotated bibliography*. London, UK: Overseas Development Institute. Retrieved from <https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9184.pdf>
- McAuliffe, C. (2012). Graffiti or street art? Negotiating the moral geographies of the creative city. *Journal of Urban Affairs*, 34, 189–206. <https://doi.org/10.1111/j.1467-9906.2012.00610.x>
- Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). New York: Cambridge University Press.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85(3), 207. <https://doi.org/10.1037/0033-295X.85.3.207>
- Morey, R. D., Rouder, J. N., & Jamil, T. (2015). BayesFactor 0.9.11-1. Comprehensive R archive network. Retrieved from <http://cran.r-project.org/web/packages/BayesFactor/index.html>. Accessed February 1, 2018.
- Newman, G. E., & Bloom, P. (2012). Art and authenticity: The importance of originals in judgments of value. *Journal of Experimental Psychology: General*, 141(3), 558. <https://doi.org/10.1037/a0026035>
- Oforu, E. K., Chambers, M. K., Chen, J. M., & Hehman, E. (2019). Same-sex marriage legalization associated with reduced implicit and explicit antigay bias. *Proceedings of the National Academy of Sciences of the United States of America*, 116(18), 8846–8851. <https://doi.org/10.1073/pnas.1806000116>
- Parker, K., Horowitz, J. M., & Stepler, R. (2017). On gender differences, no consensus on nature vs. nurture. Retrieved from Pew Research Center's Social & Demographic Trends Project website: <https://www.pewresearch.org/social-demographic-trends/2017/08/01/gender-differences-nature-vs-nurture/>

- pewsocialtrends.org/2017/12/05/on-gender-differences-no-consensus-on-nature-vs-nurture/. Accessed December 12, 2018.
- Pinal, G. D., & Reuter, K. (2017). Dual character concepts in social cognition: Commitments and the normative dimension of conceptual representation. *Cognitive Science*, 41(S3), 477–501. <https://doi.org/10.1111/cogs.12456>
- Pinker, S. (2008). The moral instinct. *The New York Times*. Retrieved from <https://www.nytimes.com/2008/01/13/magazine/13Psychology-t.html>. Accessed December 12, 2018.
- Prentice, D. A., & Miller, D. T. (2007). Psychological essentialism of human categories. *Current Directions in Psychological Science*, 16(4), 202–206. <https://doi.org/10.1111/j.1467-8721.2007.00504.x>
- Putnam, H. (1973). Meaning and reference. *Journal of Philosophy*, 70(19), 699–711. <https://doi.org/10.2307/2025079>
- Reuter, K. (2019). Dual character concepts. *Philosophy Compass*, 14(1), e12557. <https://doi.org/10.1111/phc3.12557>
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4), 573–605. [https://doi.org/10.1016/0010-0285\(75\)90024-9](https://doi.org/10.1016/0010-0285(75)90024-9)
- Rouder, J. N., & Morey, R. D. (2012). Default Bayes factors for model selection in regression. *Multivariate Behavioral Research*, 47(6), 877–903. <https://doi.org/10.1080/00273171.2012.734737>
- Ruble, D. N., Taylor, L. J., Cyphers, L., Greulich, F. K., Lurye, L. E., & Shrout, P. E. (2007). The role of gender constancy in early gender development. *Child Development*, 78(4), 1121–1136. <https://doi.org/10.1111/j.1467-8624.2007.01056.x>
- Smiler, A. P., & Gelman, S. A. (2008). Determinants of gender essentialism in college students. *Sex Roles*, 58(11–12), 864–874. <https://doi.org/10.1007/s11199-008-9402-x>
- Tankard, M. E., & Paluck, E. L. (2017). The effect of a supreme court decision regarding gay marriage on social norms and personal attitudes. *Psychological Science*, 28(9), 1334–1344. <https://doi.org/10.1177/0956797617709594>
- Tate, C. C., Youssef, C. P., & Bettergarcia, J. N. (2014). Integrating the study of transgender spectrum and cisgender experiences of self-categorization from a personality perspective. *Review of General Psychology*, 18(4), 302–312. <https://doi.org/10.1037/gpr0000019>
- Tobia, K. P., Newman, G. E., & Knobe, J. (2019). Water is and is not H<sub>2</sub>O. *Mind & Language*, 35(2), 183–208. <https://doi.org/10.1111/mila.12234>
- U.S. Bureau of Labor. (2019). Women in the labor force: A databook. Retrieved from <https://www.bls.gov/opub/reports/womens-databook/2019/>. Accessed October 9, 2020.
- Vandello, J. A., & Bosson, J. K. (2013). Hard won and easily lost: A review and synthesis of theory and research on precarious manhood. *Psychology of Men & Masculinity*, 14(2), 101. <https://doi.org/10.1037/a0029826>
- Vandello, J. A., Bosson, J. K., Cohen, D., Burnaford, R. M., & Weaver, J. R. (2008). Precarious manhood. *Journal of Personality and Social Psychology*, 95(6), 1325–1339. <https://doi.org/10.1037/a0012453>
- Weisman, K., Dweck, C. S., & Markman, E. M. (2017). Rethinking people's conceptions of mental life. *Proceedings of the National Academy of Sciences of the United States of America*, 114(43), 11374–11379. <https://doi.org/10.1073/pnas.1704347114>
- White, A., Storms, G., Malt, B. C., & Verheyen, S. (2018). Mind the generation gap: Differences between young and old in everyday lexical categories. *Journal of Memory and Language*, 98, 12–25. <https://doi.org/10.1016/j.jml.2017.09.001>
- Wiktor, V., & Jonkers, H. M. (2011). Quantification of crack-healing in novel bacteria-based self-healing concrete. *Cement and Concrete Composites*, 33(7), 763–770. <https://doi.org/10.1016/j.cemconcomp.2011.03.012>
- Wittgenstein, L. (1953). *Philosophical investigations*. Oxford, UK: Basil Blackwell.

### **Supporting Information**

Additional supporting information may be found online in the Supporting Information section at the end of the article:

**Appendix S1.** Details of Bayes Factor Analyses and Specifics of Experiment 4.