Supplementary Materials

Moral values reveal the causality implicit in verb meaning

A: Notes on Methodological Differences between Study 1 and the Replication Datasets

Replication Dataset 1. As in Study 1, participants completed a block of the implicit causality task and then entered a separate block; in this separate block, the events were re-presented with the pronoun they had selected in the first block (e.g., “Max verbed Jess because he…) and an empty text box appeared after the pronoun with the following prompt: “please finish the sentence.” Participants typed a completion to each sentence. They also filled out measures of demographics, victim stigmatization and sensitivity, and moral values as in Study 1 (data and materials available at https://github.com/ BLINDED FORREVIEW).

Replication Dataset 2. As in the previous studies, participants completed the implicit causality task, but the task here involved a total of 48 verbs. See Table 1 in the main text for the complete list of verbs. Participants also filled out measures of demographics, victim stigmatization and sensitivity, and moral values as in the other studies. They also completed the Ambivalent Sexism Inventory (Glick & Fiske, 1996); however, the present analyses did not involve the ASI. Data and materials are available at https://github.com/ BLINDED FORREVIEW.

B: Gender Condition and Implicit Causality Object-Bias from Study 1 and the Replication Datasets

To investigate whether the gender condition (male-verbed-female *versus* female-verbed-male) predicted the implicit causality object-bias for harm/force verbs, we first computed a generalized linear mixed-effects regression model in which verb type (harm/force (coded as 0) *versus* neutral filler (coded as 1)) and gender condition (male-verbed-female (coded as 0) *versus* female-verbed-male (coded as 1)) were included as fixed predictors of the propensity to select the object (coded as 1) *relative to* the subject (coded as 0) as the referent. Participant and verb were included as crossed random effects with random intercepts only. There was a significant interaction between verb type and gender condition in Study 1 and in the Replication Datasets (see **Supplementary Table 1** for full results from these models). To further interrogate these significant interaction effects, follow-up generalized linear mixed-effects models were computed for harm/force verbs and for neutral filler verbs, taken separately.

Supplementary Table 1. *The results of two generalized linear mixed-effects regression models—each with verb type and gender condition as predictors of selecting the object versus the subject as the referent.*

	b	***SE***	***Z***	***p***	**95% CI**	**Odds ratio**
Study 1						
Verb Type	1.67	.40	4.19	< .0001	[.89, 2.45]	5.30
Gender Condition	.53	.11	4.89	< .0001	[.32, .75]	1.70
Verb Type x Gender Condition	-.99	.10	-10.29	< .0001	[-1.18, -.80]	.37
Replication Dataset 1						
Verb Type	1.52	.39	3.95	< .0001	[.77, 2.28]	4.59
Gender Condition	.71	.15	4.77	< .0001	[.42, 1.00]	2.03
Verb Type x Gender Condition	-.78	.13	-5.99	< .0001	[-1.03, -.52]	.46
Replication Dataset 2						
Verb Type	2.06	.45	4.58	< .0001	[1.18, 2.94]	7.82
Gender Condition	.80	.08	10.09	< .0001	[.65, .96]	2.23
Verb Type x Gender Condition	-1.14	.08	-.15.12	< .0001	[-1.29, -.99]	.32

Note. Study 1 (*N* = 459), Replication Dataset (*N* = 249), Replication Dataset 2 (*N* = 788). All 95% CIs are for the beta-estimates.

For harm/force verbs, a generalized linear mixed-effects regression model was computed for which gender condition was included as the fixed predictor of the propensity to select the object (coded as 1) *relative to* the subject (coded as 0) as the referent (participant and verb were both included as random effects with random intercepts only). In Study 1, this analysis yielded a significant relationship between gender condition and the likelihood of selecting the object versus the subject as the referent (b = .63, SE = .13, Z = 4.98, p < .0001, odds ratio = 1.87, 95% CI = [.38, .87]). We obtained the same pattern of results in Replication Dataset 1 (b = .80, SE = .18, Z = 4.35, p < .0001, odds ratio = 2.23, 95% CI = [.44, 1.16]) and in Replication Dataset 2 (b = .90, SE = .10, Z = 9.42, p = .0001, odds ratio = 2.47, 95% CI = [.72, 1.09]). In all datasets, when women harmed/forced men, participants were more likely to select the object than the subject as the referent (i.e., there was a more pronounced object-bias).

For neutral filler verbs, there was a significant relationship between gender condition and the propensity for selecting the object over the subject as the referent in Study 1 (b = -.43, SE = .11, Z = -3.80, p = .0001, odds ratio = .65, 95% CI = [-.66, -.21]). However, in Replication Dataset 1, there was no significant relationship between gender condition and the propensity for selecting the object over the subject as the referent (b = -.04, SE = .14, Z = -.28, p = .78, odds ratio = .96, 95% CI = [-.32, .24]). Finally, in Replication Dataset 2, there was a significant relationship between gender condition and the propensity for selecting the object over the subject as the referent (b = -.32, SE = .08, Z = -4.10, p < .0001, odds ratio = .72, 95% CI = [-.48, -.17]). The significant relationships identified for neutral filler verbs were in the opposite direction of the relationships for harm/force verbs. In sum, moral values aside, participants generally were more likely to select men for harm/force events in the implicit causality task regardless of whether they were the subject (the “perpetrator” of harm/force) or the object (the “victim” of harm/force).

Importantly, when gender condition was added to the generalized linear mixed-effects models that already included binding values as a predictor of the propensity to select the object relative to the subject as the referent, participants higher in binding values were still significantly more likely to select the object over the subject as referent for harm/force verbs in Study 1 (b = .41, SE = .06, Z = 6.76, p < .0001, odds ratio = 1.50, 95% CI = [.29, .52]), Replication Dataset 1 (b = .50, SE = .10, Z = 4.95, p < .0001, odds ratio = 1.64, 95% CI = [.30, .69]), and Replication Dataset 2 (b = .22, SE = .05, Z = 4.13, p < .0001, odds ratio = 1.25, 95% CI = [.12, .32]).

C: Demographic Controls and Implicit Causality Object-Bias

Having found that for harm/force verbs, binding values significantly predict the likelihood of selecting the object over the subject as referent ("object-bias"), we next expanded the generalized linear mixed-effects regression models to include political orientation, gender, and religiosity as additional fixed predictors along with binding values. Given that prior work has identified relationships between binding values and political orientation, gender, and religiosity (Graham et al., 2011), we wanted to ensure that binding values predicted the implicit causality object-bias above and beyond these other variables. Specifically, binding values, political orientation, gender (0 = male, 1 = female), and religiosity were included as fixed predictors of the propensity to select the object (coded as 1) *relative to* the subject (coded as 0) as the referent for the harm/force verbs. Full results for these models for Study 1 and the replication datasets are depicted in **Supplementary Table 2**. Binding values remained a consistent, significant predictor of the object-bias in all three datasets after statistically controlling for political orientation, gender, and religiosity.

Supplementary Table 2. *The results of two generalized linear mixed-effects regression models—each with binding values, political orientation, gender, and religiosity as predictors of the propensity to select the object as the referent for harm and force events.*

	b	*SE*	*Z*	*P*	Odds ratio	95% CI
Study 1						
Binding Values	.21	.08	2.73	.006	1.23	[.06, .36]
Political Orientation	-.07	.04	-1.59	.111	.94	[-.15, .01]
Gender	-.59	.12	-4.92	< .0001	.56	[-.82, -.35]
Religiosity	.09	.03	2.69	.007	1.09	[.02, .15]
Replication Dataset 1						
Binding Values	.53	.13	4.08	< .0001	1.70	[.28, .79]
Political Orientation	.08	.06	1.31	.190	1.09	[-.04, .21]
Gender	-.35	.18	-1.92	.055	.70	[-.71, .01]
Religiosity	.00	.05	-.08	.938	1.00	[-.11, .10]
Replication Dataset 2						
Binding Values	.25	.07	3.69	.0002	1.28	[.12, .38]
Political Orientation	-.04	.04	-1.01	.31	.96	[-.11, .03]
Gender	-.24	.11	-2.26	.024	.79	[-.44, -.03]
Religiosity	-.05	.03	-1.69	.090	.95	[-.10, .01]

Note. Study 1 (N = 459), Replication Dataset 1 (N = 249), Replication Dataset 2 (N = 788). All 95% CIs are for the beta-estimates.

D: Individualizing Values and Implicit Causality Object-Bias

We investigated whether individualizing values also predict a subject- or object-bias in the implicit causality task. Niemi and Young (2016) found that individualizing values are positively associated with perpetrator blame. This association was notably weaker than the associations between binding values and judgments of victims as blameworthy and responsible (Niemi & Young, 2016). Therefore, we did not have strong expectations regarding the implicit

causality behavior of participants high in individualizing values. Nevertheless, increased selection of the subject for harm/force verbs would be consistent with the prior findings of increased perpetrator blame. To investigate this possibility, generalized linear mixed-effects regression models were computed in which verb type (harm/force (coded as 0) *versus* neutral filler (coded as 1)) and individualizing values were included as fixed predictors of the propensity to select the object (coded as 1) *relative to* the subject (coded as 0) as the referent.

In Study 1, there was no significant effect of individualizing values (b = -.02, SE = .08, Z = -.27, p = .79, odds ratio = .98, 95% CI = [-.18, .13]), no significant main effect of verb type (b = .88, SE = .52, Z = 1.72, p = .086, odds ratio = 2.42, 95% CI = [-.13, 1.90]), and no significant interaction (b = .06, SE = .07, Z = .85, p = .40, odds ratio = 1.06, 95% CI = [-.08, .20]). There was a significant interaction in Replication Dataset 1 (b = .23, SE = .09, Z = 2.44, p = .015, odds ratio = 1.25, 95% CI = [.04, .41]), but there were no significant main effects (ps > .50). Despite this significant interaction effect, there was still no effect of individualizing values on the propensity to select the object *relative to* the subject as the referent for the subset of harm/force verbs (p > .05) or for the subset of neutral filler verbs (p > .05), when these verb types were included in separate follow-up models. In Replication Dataset 2, there were no significant main effects, and there was no significant interaction (all ps > .05). Therefore, we found no clear, consistent relationship between individualizing values and the object bias across the three datasets.