**MLM Results**

The model specified in Equation 1 contains effects of the forward strength of the cue as well as the experimental group an individual was assigned to. In addition, participants and cues were allowed to vary in their individual intercepts. Results of the fixed effects of the model indicate that there is a significant interaction between forward strength and group. When comparing the control and experimental groups we find a significant difference in the influence of forward strength (p=.03), but no differences between the control and the matched group (p=.15), or the matched and control group (p=.52). The control group showed the largest influence of forward strength on judgment (β=0.28), followed by the experimental group (β=0. 23), and finally the matched group (β=0. 20). Results of the random effects of the model indicate that there is substantial variability in overall judgment by participant (SD=9.85), and that there is also a notable amount of variability in judgments based upon the random intercept for cue word (SD=4.49). Additionally, we examined models that included a random intercept for the target word, that removed the random intercept for the cue word, and that removed the interaction effect. All of these models indicated degraded fit and were therefore not included.

The model specified in Equation 2 contains effects of the experimental norms as well as the experimental group (only considering the experimental and matched groups) an individual was assigned to. In addition, participants and cues were allowed to vary in their individual intercepts. Results of the fixed effects of the models indicated a significant difference between the experimental and the matched group when examining the relationship between the experimental norms and judgment (p<.001), such that experimental norms were more influential for those in the experimental condition (β =.54) as compared to the matched condition (β=.21). Results of the random effects of the model indicate variability both in the overall judgment by participants (SD=10.68), as well as the random intercept for cue word (SD=4.27).