

The Role of Experimenter Belief in Social Priming



Thandiwe S. E. Gilder¹ and Erin A. Heerey^{1,2}

¹School of Psychology, Bangor University, and ²Department of Psychology, Western University

Psychological Science
2018, Vol. 29(3) 403–417
© The Author(s) 2018
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797617737128
www.psychologicalscience.org/PS



Abstract

Research suggests that stimuli that prime social concepts can fundamentally alter people's behavior. However, most researchers who conduct priming studies fail to explicitly report double-blind procedures. Because experimenter expectations may influence participant behavior, we asked whether a short pre-experiment interaction between participants and experimenters would contribute to priming effects when experimenters were not blind to participant condition. An initial double-blind experiment failed to demonstrate the expected effects of a social prime on executive cognition. To determine whether double-blind procedures caused this result, we independently manipulated participants' exposure to a prime and experimenters' belief about which prime participants received. Across four experiments, we found that experimenter belief, rather than prime condition, altered participant behavior. Experimenter belief also altered participants' perceptions of their experimenter, suggesting that differences in experimenter behavior across conditions caused the effect. Findings reinforce double-blind designs as experimental best practice and suggest that people's prior beliefs have important consequences for shaping behavior with an interaction partner.

Keywords

social power, priming, experimenter effects, open data, preregistered

Received 7/31/15; Revision accepted 9/19/17

Priming, the act of influencing another's behavior via indirect cues, is a common experimental manipulation in social psychology (e.g., Anderson & Galinsky, 2006; Dreisbach & Boettcher, 2011; Fan & Gruenfeld, 1998; Galinsky, Magee, Gruenfeld, Whitson, & Liljenquist, 2008; Overbeck & Park, 2006). In many social psychological priming paradigms, an experimenter asks participants to perform a task that *primes*, or activates, a particular concept, such as age or social power. The prime's effect is then examined in a subsequent task. Although participants appear to be unaware of the relation between the prime and target task, the prime nonetheless affects target-task performance.

Although priming appears to be one way of influencing behavior, the subtle social cues people exchange in face-to-face interactions also have powerful effects (Rosenthal, 1994). Indeed, the beliefs and stereotypes that people bring to interactions shape both the behaviors they produce (Wheeler & Petty, 2001) and their interaction partners' responses (Herr, Sherman, & Fazio, 1983). For example, when one interaction partner holds

a stereotype about another, that partner is likely to behave more stereotypically, even when the belief holder does not intend to transmit the stereotype (Snyder & Stukas, 1999).

In research settings, experimenters' expectations may have the insidious effect of confounding task results. Indeed, research shows that when experimenters are motivated to find significant effects, they are more likely to do so (Sheldrake, 1998). Because expectations can be quite powerful, they may bias experimenter behavior when experimenters are aware of both the study hypotheses and participants' conditions (Rosenthal, 1994; but see Barber, 1978). Changes in experimenter behavior may subsequently cause changes in participant behavior, independently of experimental manipulations.

Corresponding Author:

Erin A. Heerey, Department of Psychology, Western University, Social Sciences Centre, Room 7418, London, Ontario, Canada N6A 5C2
E-mail: eheerey@uwo.ca

Unfortunately, in many articles in the priming literature, double-blind experimental designs are not explicitly described (e.g., Galinsky, Magee, Inesi, & Gruenfeld, 2006). This may be problematic for interpreting results. Indeed, authors of several recent empirical articles have independently reported failures to replicate findings from the priming literature under double-blind conditions (e.g., Doyen, Klein, Pichon, & Cleeremans, 2012; Harris, Coburn, Rohrer, & Pashler, 2013; Pashler, Coburn, & Harris, 2012; Shanks et al., 2013), and research suggests that failure to employ double-blind designs may be endemic (Klein et al., 2012).

For our experiments, we chose tasks thought to prime social power, defined as the ability to access, control, and distribute resources within a group (Keltner, Gruenfeld, & Anderson, 2003). Power primes have been the focus of much research, with results generally indicating reliable effects (Galinsky, Gruenfeld, & Magee, 2003; Galinsky et al., 2008; Magee, Galinsky, & Gruenfeld, 2007). For example, research suggests that experimentally priming high versus low social power may improve executive cognition, including more flexible attention, reasoning, and cognitive control as well as better ability to inhibit the influence of distractors (Galinsky et al., 2003; Guinote, 2007; Smith, Jostmann, Galinsky, & van Dijk, 2008; Willis, Rodríguez-Bailón, & Lupiáñez, 2011). High-power primes may also enhance abstract thinking, risk taking, approach behavior, and optimism (Anderson & Galinsky, 2006; Maner, Gailliot, Butz, & Peruche, 2007; Smith & Trope, 2006). However, none of these experimental reports explicitly describes a double-blind design. If experimenters were aware of both participants' conditions and the research hypotheses, they may have inadvertently altered their behavior on the basis of this knowledge, thereby communicating expectations to participants.

In Experiment 1, we used a computer-administered role-play task to assign participants to low-power ("employee") versus high-power ("boss") conditions in a double-blind design. Our aim was to conceptually replicate work demonstrating that high- versus low-power roles enhance the ability to inhibit distractors during target detection (Guinote, 2007). Despite robust effects of the power manipulation, we failed to find evidence of the expected power-priming effect on a flanker task.

However, we worried that our double-blinding procedure, which diverges from those used in typical experimental research in this area, might explain our failure to find the predicted results. We therefore sought to manipulate both priming condition and experimenter belief about priming condition simultaneously using a computerized version of a priming task commonly reported in the literature. In each of four independent experiments, involving 11 experimenters and a total of

824 participants, we used a common priming task to activate feelings of high or low social power while independently manipulating experimenter knowledge about participant condition and therefore about expected results. In Experiment 2, we measured word categorization speed (Smith & Trope, 2006, Experiment 1); in Experiment 3, we examined risk taking using the Columbia Card Task (Figner, Mackinlay, Wilkening, & Weber, 2009); in Experiment 4, we assessed abstract versus concrete categorizations of everyday behaviors (Smith & Trope, 2006, Experiment 2); and in Experiment 5, we studied approach behavior (Smith & Bargh, 2008, Experiment 2). In response to reviewer comments, we increased the experimental power of the final experiment and pre-registered it at the Open Science Framework.

Experiment 1

Our goal in Experiment 1 was to replicate previous work demonstrating that priming high versus low power would enhance participants' ability to ignore distractors (Guinote, 2007) in a flanker task. However, given that we planned a between-subjects design, we wanted to ensure that experimenter expectations would not bias data collection. We therefore used a computerized priming task to guarantee that the experimenter was entirely unaware of prime condition prior to debriefing participants.

Method

Participants. One hundred eighteen undergraduate psychology students participated in an experiment about "personality and cognition" in exchange for partial course credit and a small monetary bonus. We excluded 1 participant's data because of a computer failure that caused data loss on approximately 70% of trials. We also excluded 4 participants' data because they indicated suspicions about the link between the prime and target tasks. The final sample size was 113 participants (86 women; age: $M = 20.48$ years, $SD = 3.85$). We selected sample sizes a priori on the basis of a power analysis (two-tailed $\alpha = .05$, effect size $d = 0.70$, experimental power = .80) using typical reported effect sizes (e.g., Smith & Bargh, 2008; Smith & Trope, 2006). Participants gave written consent before participating and were fully debriefed after experiment completion. Bangor University's ethics committee approved all procedures.

Experimenter. One female experimenter (T. S. E. G.) completed all the data collection for this experiment as part of a doctoral thesis. The experimenter had read and discussed the power and executive-cognition literature with several collaborators. The experimenter believed that

she was extending findings in the power-priming literature to a flanker task and fully expected to find priming effects.

Priming task. We used a strong explicit power manipulation in which participants were assigned to high-power (“boss”), low-power (“employee”), or equal-status (control) groups for a computerized role-play task. Participants gave their consent and were instructed in pairs in order to give the impression that they would be working together in the task (in reality, all participants completed the task individually). They were then shown to adjacent rooms for the experimental procedure. After this instruction stage of the experiment, the experimenter had no further contact with participants until debriefing.

The computer randomly assigned each participant to one of two power-related roles (boss, $n = 37$, or employee, $n = 38$) for a target-detection game. Participants believed that they were working with the partner to earn bonus money in the game. A third group of participants was assigned to a cooperative control condition ($n = 38$). Because the computer assigned participants to priming conditions and administered task instructions accordingly, the experimenter was blind to condition until the debriefing phase of the experiment.

Although all participants completed the same game, the instructions differed depending on computer-assigned roles. Participants were told that their primary task was to press a key whenever they detected a target (colored square) on the left side of the screen (see the Supplemental Material available online for full details). “Bosses” were told that, as an added responsibility of their role, they should also detect and respond to targets on the right side of the screen. Employees were told that the boss had assigned them this same duty. Participants in the cooperative condition believed that they were working as a team and that both partners would respond to both left and right targets. Regardless of actual performance, participants learned that together they had earned £4.98. Bosses then assigned any amount of this bonus to their employees, retaining the remainder for themselves. On average, bosses in the experiment behaved relatively fairly, assigning 43.98% ($SD = 17.75\%$) of the total bonus to their employees. To emphasize the power differential, however, we told employees that they had been allocated 35% of the bonus. In the cooperative condition, participants were told at the task outset that they would each receive 50% of the bonus.

Following the power induction, participants completed a four-item questionnaire to measure their sense of fairness about the task (“To what extent do you feel like the workload division was fair?” and “To what extent do you feel like the bonus money was divided fairly?”), effort expended (“To what extent did you feel

like you performed the task to the best of your ability?”), and power (“To what extent did you feel powerful or in control in the task?”). These questions served as a manipulation check.

Target task. To assess power-related differences in cognitive and attentional control, we then had the participants complete a flanker task (Eriksen & Eriksen, 1974). Participants made speeded left or right button presses to indicate the direction of a central target arrow. A pair of leftward- or rightward-pointing arrows served as distractors. Trials began with a fixation cross for 500 ms, followed by a target arrow (50% pointed left) surrounded by distractor arrows pointing in either the same (congruent; 50% of trials) or the opposite (incongruent) direction. The target-flanker display remained visible for 500 ms before being replaced by a blank screen until the response. Participants then saw feedback about whether they were correct (1,000 ms). They completed three blocks of 60 randomly ordered trials each. At the end of the session, the experimenter fully debriefed participants and probed them for suspicion. All participants received the same monetary bonus (£5). The experimental protocol was fully automatized using E-Prime software (Version 1.2; Schneider, Eschman, & Zuccolotto, 2001).

Data analysis. We calculated the proportion of correct trials and the mean reaction times (RTs; excluding trials with incorrect responses) for congruent and incongruent trials as a measure of the flanker effect. Because we consider the absence of an effect to be equally as important as its presence, we examined these data using Bayesian analyses of variance (ANOVAs) with power condition (high, low, control) as the between-subjects variable. In Bayesian analysis, the presence and absence of an effect are evaluated with different models. Prior probability distributions for the coefficients under each model are specified. This procedure allowed us to calculate each model’s marginal likelihood given the observed data. The ratio of the two models’ marginal likelihoods is the Bayes factor. For model comparison, we report the Bayes factor (BF_{10}), the ratio of the probability of the observed data under the alternate model to the probability under the null model. Note that the Bayes factor automatically penalizes for model complexity, such that in the absence of any effect, the evidence favors the simpler over the more complex model. A BF_{10} greater than 1 indicates that the evidence favors the alternate model, and a BF_{10} lower than 1 suggests that the evidence favors the null model. BF_{10} s ranging from 3 to 20 are considered positive evidence in favor of the alternate model, whereas BF_{10} s ranging from 0.33 to 0.10 constitute moderate evidence in favor of the null model (see Jarosz & Wiley, 2014). Note that we report the BF_{01} (the ratio of the probability of the observed data

under the null model to the probability under the alternate model) when evidence appeared to favor the null model. We gave each of the models (e.g., null, prime condition) an equal (uninformative) prior probability. Traditional ANOVA results appear in the Supplemental Material available online. All Bayesian analyses were conducted using JASP software (Version 0.8.2, JASP Team, 2017).

Results

Manipulation check. To test the efficacy of the power prime, we conducted a set of Bayesian ANOVAs, with effort, fairness, and power ratings as dependent variables and power condition as the independent variable. With respect to self-reported effort, the results were nondiagnostic ($BF_{10} = 1.481$). That is, even though low-power participants appeared to report slightly more effort than high-power or control participants (Fig. 1a), the data did not conclusively support either the null model or an effect of prime condition. In contrast, analyses suggested that prime condition was highly effective at influencing perceptions of both task fairness ($BF_{10} = 1.868 \times 10^8$) and experienced power ($BF_{10} = 3.745 \times 10^5$). Specifically, low-power participants thought the task was less fair than did participants in the other two conditions, and they felt less powerful, especially relative to high-power participants. These results suggest that the power manipulation effectively induced feelings of high and low power. Data files for Experiment 1 are available at the Open Science Framework.

Target task. The RT (Fig. 1b¹) and accuracy (Fig. 1c) results from the flanker task show strong evidence for the presence of the typical flanker effect. Participants responded both more quickly and more accurately on trials with congruent distractors than on trials with incongruent distractors. Interestingly, when the experimenter was unaware of the power condition to which participants had been assigned, there was no indication that this effect was modulated by the prime (speed: $BF_{01} = 5.952$; accuracy: $BF_{01} = 8.333$). Thus, when the experimenter was blind to prime condition, the data were almost six times more likely for response speed (and eight times more likely for accuracy) under the null than the prime-effect hypothesis.

Discussion

Under double-blind conditions, we found no evidence that power primes affected behavior in a subsequent flanker task, despite robust effects on our manipulation check. We can think of three possibilities for this occurrence. First, the flanker task has not, to our knowledge, been used with a power prime. Nonetheless, tasks

tapping similar facets of executive cognition have shown power-priming effects (Guinote, 2007; Smith et al., 2008), and the flanker task itself may be sensitive to a social-status prime (Dreisbach & Boettcher, 2011). Second, although the power manipulation we used was based on previous role-play priming tasks, we did not actually ask participants to interact with an experimenter or each other as is typical (e.g., Galinsky et al., 2003). However, computerization of this task was necessary to ensure that the experimenter remained blind to participants' condition. Finally, our double-blind design may have played a role in the present results. We tested this idea across the next four experiments.

Experiments 2 Through 5

In these experiments, we asked whether experimenters' knowledge of participants' priming condition might influence results, independently of participants' actual task condition. To test this question, we orthogonally manipulated experimenters' belief about which prime condition each participant experienced and the actual prime condition that a participant received. In all four experiments, we used a computerized version of a power prime that has been frequently used to prime social power (e.g., Smith & Bargh, 2008; Smith & Trope, 2006). Each experiment involved an independent set of experimenters and a different target-task.

General method

In Experiments 2 through 5, we followed the same general protocol. We begin by describing this common methodology. We then describe the unique aspects and main results of each experiment separately, reserving manipulation-check data and additional experimenter-related results for a general results section. All experimental procedures were approved by the ethics committees of Bangor University (Experiments 2–4) or Western University (Experiment 5), and all participants provided fully informed consent after debriefing.

Experimenter selection and training. Experimenters were either master's-level (Experiments 2–4) or honors (Experiment 5) undergraduates who conducted the research in the context of thesis projects.² To ensure that they understood the literature and expected findings, the experimenters participated in journal clubs, in which they read and discussed a series of articles from the relevant power-priming literature. On the basis of these discussions, they developed hypotheses and selected target tasks. In all cases, they believed that they were replicating (conceptually or directly) and extending the relevant literature to account for the effects of both mood and

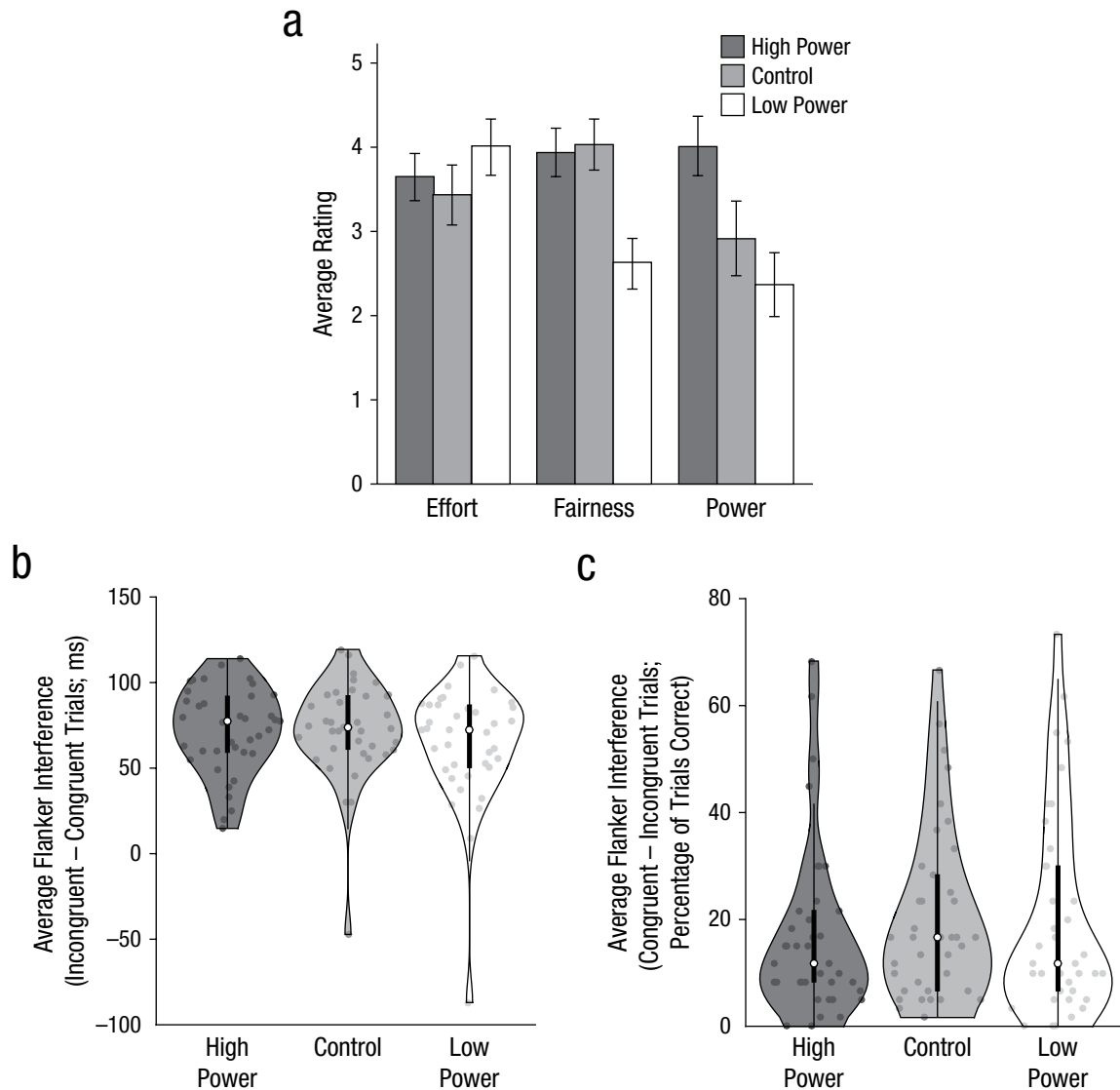


Fig. 1. Results of Experiment 1 ($N = 113$). Average ratings for perceptions of effort, fairness, and power (a) are shown as a function of condition. Error bars indicate 95% credible intervals. The violin plots (including individual data points) show the average flanker interference effects for (b) reaction time (incongruent trials – congruent trials) and (c) accuracy (congruent trials – incongruent trials) in each of the three conditions. For each condition, the white dot indicates the median, and the central box shows the interquartile range. The whiskers show the 95% credible interval for the median.

power on their target tasks. In the context of training, they each learned a script for instructing participants (see Supplementary Methods in the Supplemental Material), completed the experimental session as if they were participants, and practiced testing one another on the task.

Experimenter belief manipulation. Each experimenter independently collected data from a sample of participants. Working from a list that ostensibly assigned participant ID codes to power-prime conditions, experimenters started the computer program before each participant arrived. After entering a participant's ID, they typed "H" for high power or "L" for low power to start the task. They believed that this

procedure caused the computer to administer the high- and low-power primes. Unbeknownst to experimenters, only half the participants completed the priming condition to which the experimenter "assigned" them. In these cases, the experimenter's beliefs about the prime condition and the actual prime condition were congruent, as in past research. The remaining participants completed the opposite condition to which the experimenter believed they had been assigned, meaning that the actual prime condition differed from the experimenter's belief about it.

Experimenters tested participants individually, obtained their consent, and instructed them using a script (see the Supplemental Material). They also

answered any questions a participant asked. This procedure took about 5 min. Once participants began the computerized portion of the experimental session, they had no further contact with experimenters until debriefing.

Throughout the data collection phase of the experiment, experimenters remained blind to this manipulation. Therefore, they had knowledge only about the condition they believed participants to have completed and the results expected on the basis of that belief. We fully debriefed experimenters at the completion of data collection, and all experimenters provided informed consent for their data to be reported in this article. None reported any suspicion about the manipulation.

Power-priming task. The cover story maintained that that the experiments involved unrelated tasks and that we wanted to control for individual differences in participants' moods in our analyses. Therefore, participants were told that they would complete a baseline mood measure before each of the tasks. Consistent with this story, the computer administered the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) before both prime and target tasks, with a randomized word order. We also embedded five power-related words into the PANAS at random points ("powerless," "unimportant," "dominant," "self-assured," and "influential"; the first two of these were reverse-scored, and the words appeared in random order). These data allowed us to measure changes in feelings of power from pre- to postmanipulation and served as a manipulation check for the power prime. Analysis of Cronbach's alpha showed that the set of items had moderate to good reliability ($\alpha = .729$), and a principal component analysis confirmed that the items loaded onto a single factor with loadings greater than .638. Embedding these words within the PANAS helped to conceal the nature of the experimental manipulation. Participants rated the degree to which they felt each item "right now" on a 100-point visual analogue scale by clicking a mouse.

After the PANAS, participants completed the power prime, a computerized version of the same 17-item scrambled-sentence priming task reported in Smith and Trope (2006, Experiment 2). On each trial, they created grammatically correct sentences using a mouse to select and organize four of five randomly ordered words (e.g., "class," "he," "dominates," "the," and "chooses"). In the high-power condition, half of the sentences included words associated with high power (e.g., "dominates," "commands"), and in the low-power condition, half of the sentences contained words associated with low power (e.g., "subordinate," "obeys"). Participants spent as long as they liked working on each sentence and could click an "undo" button if they made a mistake.

Task items and word orders were identical to those in previous research (Smith & Bargh, 2008; Smith & Trope, 2006). After the second PANAS, participants completed the target task associated with their experiment.

Finally, we wanted to assess whether experimenters' expectations altered the impressions they made on participants. To achieve this, we programmed the computer to ask participants to rate the experimenter after the target task. Participants responded to the prompt, "To what extent do you think the experimenter is:" and rated the experimenter on the following adjectives: "attractive," "competent," "friendly," and "trustworthy." Responses were made on a 7-point Likert scale (1 = *not at all*, 7 = *extremely*). Experimenters were unaware that participants made these ratings. The experimental protocol was fully automatized and presented using E-Prime software (Version 1.2 in Experiments 2–4 and Version 2 in Experiment 5). All participants were tested individually. At the end of the session, the experimenter returned to the room to debrief and probe each participant for suspicion about the purpose of the experiment and the relation between the tasks using a funnel-debriefing procedure (Bargh & Chartrand, 2014).

Experiment 2 method

Participants. One hundred sixteen psychology undergraduates participated in an experiment about "cognition and mood" in exchange for partial course credit. We excluded 5 participants' data because of poor English fluency (they all needed the aid of a dictionary during the target task). The final sample therefore included data from 111 participants (77 women; age: $M = 21.64$ years, $SD = 4.44$). Our aim in selecting this sample size was to balance experimental power, assuming a two-tailed α of .05, an effect size d of 0.70 (e.g., Smith & Trope, 2006), and power of .80, as well as feasibility of project completion within the allotted time.

One male and one female experimenter collected data for this experiment. They believed that the project was a conceptual extension of Smith and Trope's (2006, Experiment 1) findings on the effects of power and abstract thinking. They thought they were extending previous findings by examining participants' RTs on a word categorization task (unreported in Smith and Trope's article) and changing the priming task from a prompted writing task to our computerized scrambled-sentences task.

Target task. To measure the influence of prime and experimenter expectation on abstract thinking ability, we asked participants to complete an English-language version of the word categorization task reported in Experiment 1 of Smith and Trope's (2006) article. We used the same categories and exemplars as Smith and Trope (vehicles, furniture,

and clothing), presented in random order. On each trial, participants saw the category name at the top of the screen with a category exemplar below it. They rated how well they thought the exemplar fit into the category (using a 10-point scale from 0, *item does not belong in this category*, to 9, *item definitely belongs in this category*; see Fig. S1A in the Supplemental Material, for an example). Participants responded as quickly as possible to a total of 18 exemplars in each category (e.g., “vehicle”); 6 of these exemplars were weak (e.g., “feet”), 6 were moderate (e.g., “helicopter”), and 6 were strong (e.g., “car”). The first item from a category was always a strong exemplar, and the remaining items appeared in random order. The experimenters believed that participants receiving the high-power prime would classify category exemplars more quickly. The dependent variable for this task was mean RT across all trials. We analyzed these data using Bayesian ANOVAs with experimenter belief (high, low) and prime condition (high, low) as between-subjects factors.

Experiment 2 results

As Figure 2a shows, the mean RTs for the two priming conditions appeared to be similar. Accordingly, Bayesian analysis suggested that the data were almost 5 times more likely under the null model than under the priming-effect model ($BF_{01} = 4.926$). In contrast, analyses showed positive evidence in favor of the experimenter-effect model, relative to the null model ($BF_{10} = 3.179$). Full Bayesian results for all tested models (e.g., the interaction) appear in the Supplemental Material. These results provide moderate evidence for a model that included an experimenter effect and suggest that the null model was superior to the model allowing for a priming effect. We also note that we failed to find evidence of priming effects on actual categorization ratings, contrary to Smith and Trope (2006; see the Supplemental Material).

Experiment 3 method

Participants. One hundred ten undergraduate psychology students (66 women; age: $M = 21.18$ years, $SD = 3.71$ years) participated in an experiment about “motivation and mood” in exchange for partial course credit and a small performance-based monetary bonus. One male and one female experimenter collected the data. The experimenters believed that the project was a conceptual replication of Maner et al.’s (2007) study, in which participants primed with high power took more risks.

Target task. To assess risk-taking behavior, we asked participants to complete the “hot” version of the Columbia Card Task (CCT; see Figner et al., 2009). The CCT is a

sequential risk-taking task, in which participants make a series of selections from a field of cards (Fig. S1B in the Supplemental Material). Each field contains mostly “gain” cards (yellow happy faces), for which participants earn points, and up to three “loss” cards (green unhappy faces) that lead to punishment if uncovered. Participants click on cards one at a time to reveal outcomes. If the click reveals a gain card, participants earn points and may choose another card. If it reveals a punishment card, the trial immediately ends, and the loss is deducted from the trial earnings. As long as no loss card has been revealed, participants may stop a trial at any time (even if they have not selected any cards). Because each selected gain card increases the ratio of loss:gain cards, each click is more risky than the previous (see Supplementary Methods in the Supplemental Material for additional detail). Participants completed 27 trials of the task and received a small cash bonus equal to the number of points they earned on three randomly selected trials at the end of the experiment.

As a measure of risk taking, we used the average number of cards selected per trial. Because the loss cards were randomly distributed in each deck, occasionally the trial ended during an early click. To ensure that these random occurrences did not influence our dependent measure, we used only trials in which participants stopped voluntarily (Figner et al., 2009).

Experiment 3 results

Although experimenters expected participants primed with high power to engage in more risk taking, the evidence did not strongly suggest either the null model or the priming-effects model, $BF_{01} = 2.674$. However, there appeared to be a strong influence of experimenter belief on participants’ risk-taking behavior (Fig. 2b). A Bayesian ANOVA indicated that the data were 25 times more likely under the experimenter-effects model than under the null model ($BF_{10} = 25.088$; see Fig. 2b). Full Bayesian results for all tested models appear in the Supplemental Material.

Experiment 4 method

Participants. One hundred eighty-one undergraduate psychology students participated in an experiment about “cognition and mood” in exchange for partial course credit. We excluded data from 2 female participants, 1 for extremely fast responding throughout the task (all RTs < 200 ms, suggesting that she had not read the items), and 1 for indicating suspicion about the prime’s relation to the target task. The final sample included 179 participants (151 women; age: $M = 20.26$ years, $SD = 3.47$) and three female experimenters.

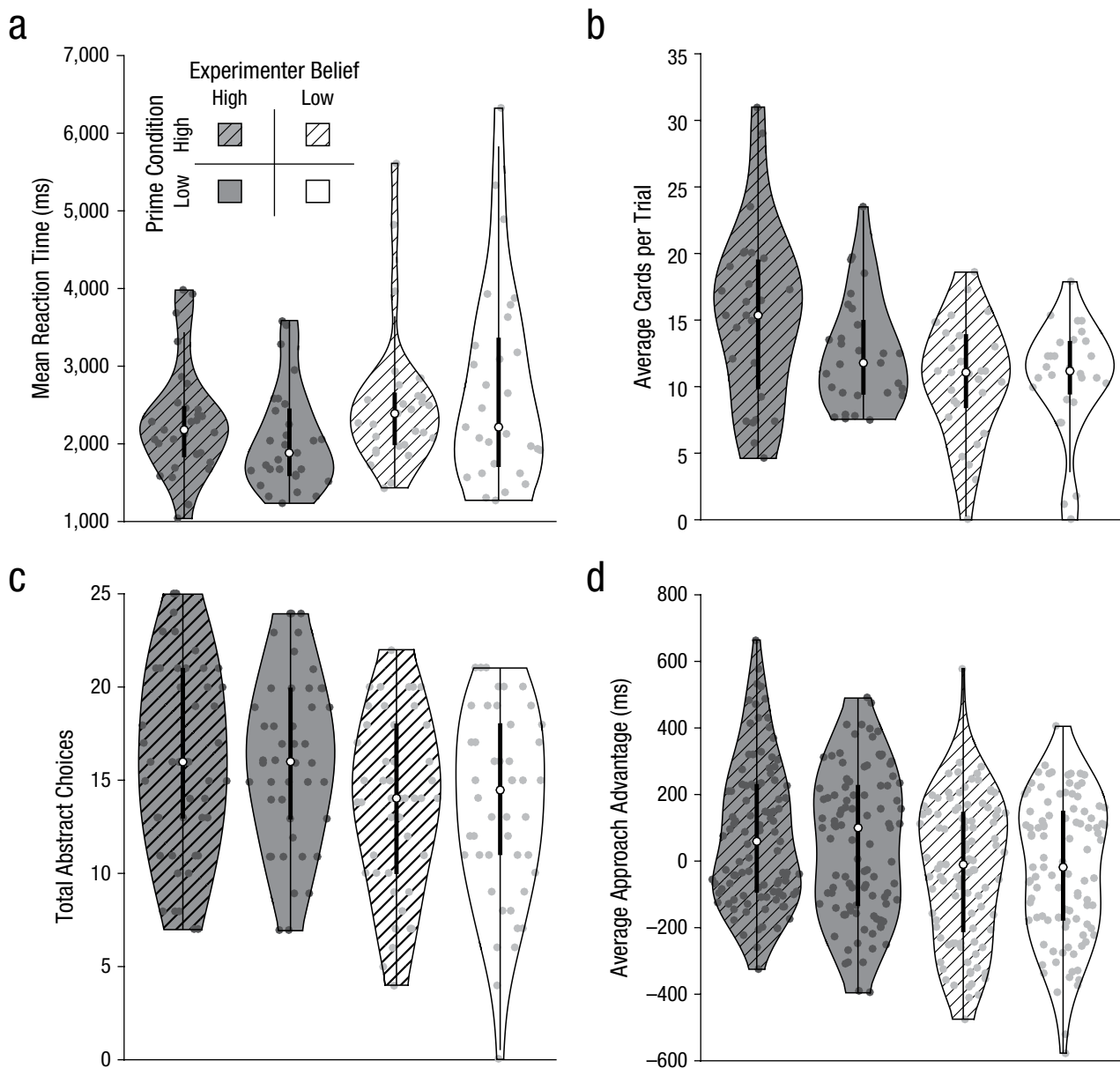


Fig. 2. Results of Experiments 2 through 5. The violin plots (including individual data points) show (a) mean reaction time for the exemplar classification task in Experiment 2 ($N = 111$); (b) the average number of cards selected per trial on the Columbia Card Task in Experiment 3 ($N = 110$); (c) the total number of abstract choices on the Behavior Identification Form in Experiment 4 ($N = 179$); and (d) the average approach advantage (avoidance trials – approach trials) in the lexical decision task in Experiment 5 ($N = 400$). In each graph, results are shown separately for each combination of prime condition and experimenter belief. White dots indicate the median, and the central boxes show the interquartile range. The whiskers show the 95% credible interval for the median.

Target task. Experiment 4 was a direct replication of Smith and Trope's (2006, Experiment 2) finding on abstract categorizations of everyday behavior. Participants completed the Behavior Identification Form (Vallacher, Wegner, & Somoza, 1989), which lists 25 common behaviors, each followed by two alternative descriptors for the behavior (e.g., the behavior "reading" might be classified as "following lines of print" or "gaining knowledge"; see Fig. S1C in

the Supplemental Material). Participants chose the descriptor that best characterized each action for them. One of the descriptors was always a more abstract description of the behavior, and the other was more concrete. The dependent variable was the number of abstract classifications participants made. The experimenters predicted that participants primed with high power would make more abstract categorizations than participants primed with low power.

Experiment 4 results

The data (see Fig. S2C in the Supplemental Material) showed positive evidence favoring the null model over the model including the priming effect, $BF_{01} = 6.098$. As in the previous experiments, however, the evidence strongly supported the experimenter-belief model, relative to the null model, $BF_{10} = 20.760$. Full Bayesian results for all tested models appear in the Supplemental Material. Thus, the effects of experimenter belief appeared to be more likely than priming effects.

Experiment 5 method

Participants. Here, we enhanced our sample size to ensure adequate statistical power in response to reviewer feedback. An a priori G*Power 3.1 analysis (Faul, Erdfelder, Buchner, & Lang, 2009; $\alpha = .05$, $\eta^2 = .04$) suggested that a sample size of 400 participants was sufficient to achieve 95% power to detect a main effect of prime condition on the basis of previously reported effects (e.g., Smith & Bargh, 2008). In exchange for partial course credit, 417 undergraduate psychology students participated in an experiment about “cognition and mood.” Per reviewer suggestion, this experiment (including methodology, sample size, exclusion criteria, hypotheses, and analyses) was preregistered at the Open Science Framework prior to data collection. To ensure that the experimenter belief manipulation remained secret, we embargoed relevant aspects of the protocol until after experimenter debriefing. Following our preregistered procedures, we excluded data from 17 participants because of poor task performance ($> 20\%$ of trials affected by errors, $RTs < 250$ ms, or $RTs > a$ participant's grand mean + 3 SDs) or speaking to the experimenter during the experimental session. The final sample consisted of 400 participants (291 women; age: $M = 18.480$ years, $SD = 1.288$). There were three female experimenters and one male experimenter.

Target task. The target task was a direct replication of the lexical decision task described in Smith and Bargh (2008, Experiment 2), in which participants primed with high power were faster to engage in approach behavior. The only difference from the published task was that we switched the words from Dutch to English. In the task, participants responded to a series of centrally presented letter strings by pressing a key that moved a stick figure either toward or away from the letter string, depending on whether or not it was an English word. Each participant completed the task following one of two movement instructions. They either moved the stick figure toward words (approach direction) and away from nonwords (avoid direction) or away from words and toward

nonwords (counterbalanced across experimenter belief and prime condition).

On each trial of the task, participants viewed a central fixation cross for 2,000 ms. A stick figure then appeared, centered in either the top half (50% of trials) or bottom half of the screen. After an onset delay of 750 ms, a central letter string appeared and remained visible until participants pressed either the up or down arrow key on the keyboard (see Fig. S1D in the Supplemental Material). After the key press, the stick figure moved toward the center or edge of the screen. After 750 ms, the next trial began. Participants were told to keep their fingers on the response keys and respond as quickly and accurately as possible.

The computer measured RT from the onset of the letter string to the first key press. There were 24 trials containing English words (in a random set of 12 of these trials, the stick figure appeared above the stimulus; likewise for nonword trials) and 24 trials containing nonwords (stimuli are available at the Open Science Framework). The words were rated as medium in frequency and neutral in valence on the basis of a set of published word norms (Warriner, Kuperman, & Brysbaert, 2013) along with ratings from an independent sample of 98 local participants. Trials appeared in random order. Prior to beginning the task, participants completed 12 practice trials,³ with speed and accuracy feedback given after each. There was no feedback during the task (the E-Prime script used in data collection is available at the Open Science Framework).

Because a preliminary analysis indicated that instruction set (i.e., approach words or approach nonwords) did not moderate task results (all $ps > .406$), we collapsed across this variable, as Smith and Bargh (2008) did. To examine the effects of experimenter belief and prime condition, we calculated the *approach advantage* that participants experienced in the task by subtracting mean approach speed from mean avoidance speed (excluding trials with errors and trials in which RT was less than 250 ms or more than 3 SDs above a participant's mean). This preregistered performance index served as the dependent variable. Data analysis was fully automatized, such that it could not be influenced by experimenter expectations.

Experiment 5 results

Experimenters expected an approach advantage for participants primed with high power. As in the previous experiments, evidence suggested that the data were almost 8 times more likely under the null model than under the prime-only model, $BF_{01} = 7.813$ (see Fig. 2d), and very strongly supported the experimenter-belief-only model, relative to the null model, $BF_{10} = 537.388$.

Full Bayesian results for all tested models appear in the Supplemental Material. Thus, across all four of these experiments, results favored an experimenter-effects model relative to the null model and provided moderate evidence for the null model relative to the priming-effects model. A mini-meta-analysis of our results appears in the Supplemental Material.

General results of Experiments 2 through 5

Manipulation check. To ensure that the priming task activated power-related concepts, we used the power items hidden in the PANAS as a manipulation check. We use frequentist analyses to describe our results so readers can compare the effects of our implicit power manipulation with those in previous research reports. Bayesian results appear in the Supplemental Material.

Although priming effects are often not directly measured, reports from the power-priming literature suggest that the high-power version of the scrambled-sentences priming task induces greater feelings of power than does the low-power version. We tested whether the prime influenced feelings of power in Experiments 2 through 5 using the power-related items embedded in the PANAS. We therefore examined whether prime condition influenced postprime feelings of power using analysis of covariance models with preprime feelings of power as the covariate.

In Experiment 2, in contrast with predictions (e.g., Galinsky et al., 2003; Smith et al., 2008; Smith & Trope, 2006), the priming task did not appear to have influenced participants' feelings of power, $F(1, 108) = 0.306$, $p = .581$, $d = 0.06$, 95% confidence interval (CI) = $[-0.22, 0.35]$ (adjusted mean for high-power condition = 59.61, 95% CI = $[56.59, 62.62]$; adjusted mean for low-power condition = 58.42, 95% CI = $[55.43, 61.41]$). In Experiment 3, however, the prime condition did have a statistically significant effect on feelings of power: Participants exposed to the high-power prime felt more powerful (adjusted mean = 61.81, 95% CI = $[58.02, 65.59]$) than did those exposed to the low-power prime (adjusted mean = 56.05, 95% CI = $[52.20, 59.91]$), $F(1, 107) = 4.464$, $p = .037$, $d = 0.36$, 95% CI = $[0.00, 0.72]$. We found similar results in Experiment 4, $F(1, 176) = 5.763$, $p = .017$, $d = 0.18$, 95% CI = $[-0.03, 0.39]$ (adjusted mean for high-power condition = 56.65, 95% CI = $[54.91, 58.39]$; adjusted mean for low-power condition = 53.65, 95% CI = $[51.92, 55.38]$), and Experiment 5, $F(1, 397) = 15.580$, $p < .001$, $d = 0.20$, 95% CI = $[0.06, 0.34]$ (adjusted mean for high-power condition = 59.51, 95% CI = $[57.97, 61.05]$; adjusted mean for low-power condition = 55.12, 95% CI = $[53.58, 56.67]$). We note, however, that the

effect sizes are small, and Bayesian analyses suggest anecdotal support at best with respect to power-priming effects on the manipulation check (see the Supplemental Material). Nonetheless, with the exception of Experiment 2, these effects (and effect sizes) are similar to those that have been previously reported (e.g., Galinsky et al., 2003), suggesting that the power prime here was effective in changing feelings of power.

Experimenter effects. Although the experimenters in Experiments 2 through 5 achieved the empirical results they predicted on the basis of their beliefs about participants' prime condition, they each asserted that this knowledge had not affected their behavior when they instructed participants. How did experimenters transmit these effects? To examine this question, we asked whether participants' ratings of experimenters depended on experimenter belief. Because people's interpersonal behavior varies dramatically depending on a variety of factors (e.g., personality; Sherman, Rauthmann, Brown, Serfass, & Jones, 2015), we had no a priori hypotheses about which experimenter ratings would differ or whether they would do so consistently across the set of experimenters—only that some characteristics would differ for experimenters who produced moderate experimenter effects (as noted in pre-registration). We conducted frequentist and Bayesian ANOVAs for each experimenter using the trait ratings as dependent variables and experimenter belief as the independent variable (results appear in Table 1). For 9 of the 11 experimenters, we found statistically significant effects, although not all of these reached reportable thresholds using Bayesian models.

Detailed analysis suggests that experimenters transmitted their expectations in different ways. Generally, however, when experimenters believed that their participants were in the high-power condition, they were rated as more trustworthy, often friendlier (although some experimenters were rated as less friendly), and sometimes more attractive (see Table 1) than those in the low-power condition. There were no differences in participants' ratings of experimenter competence across the experimenter belief conditions, meaning that it is likely that experimenters presented task instructions clearly regardless of condition.

Interestingly, the two experimenters who were not rated differently on the basis of their beliefs about participants' priming conditions did not produce experimenter effects on their target tasks (see Table 1). Together, these results suggest that experimenters' prior beliefs shaped participants' target-task behavior, likely via subtle changes in experimenter behavior. The two exceptions suggest that some individuals may be less susceptible to producing experimenter effects than others.

Table 1. Individual Experimenter Effects

Trait	Belief condition		Comparison of belief conditions			
	Low power (<i>M</i>)	High power (<i>M</i>)	<i>F</i>	BF ₁₀	<i>p</i>	Cohen's <i>d</i>
Experimenter 1 (<i>N</i> = 60); experimental effect size: Cohen's <i>d</i> = 0.516, 95% CI = [−0.01, 1.04]						
Attractiveness	4.03 [3.45, 4.62]	3.63 [3.05, 4.22]	0.928	0.387	.339	−0.253 [−0.79, 0.36]
Competence	5.77 [5.32, 6.21]	5.73 [5.29, 6.18]	0.011	0.264	.916	−0.034 [−0.51, 0.35]
Friendliness	6.17 [5.73, 6.61]	5.27 [4.83, 5.71]	8.386	7.799	.005	−0.760 [−1.30, −0.48]
Trustworthiness	5.50 [5.02, 5.98]	5.30 [4.82, 5.78]	0.342	0.303	.561	−0.154 [−0.63, 0.32]
Experimenter 2 (<i>N</i> = 51); experimental effect size: Cohen's <i>d</i> = 0.486, 95% CI = [−0.09, 1.06]						
Attractiveness	4.07 [3.41, 4.74]	4.54 [3.83, 5.25]	0.934	0.412	.339	0.277 [−0.37, 0.96]
Competence	5.67 [5.19, 6.14]	5.08 [4.58, 5.59]	2.869	0.900	.097	−0.490 [−1.00, −0.05]
Friendliness	5.78 [5.32, 6.23]	4.71 [4.23, 5.19]	10.530	16.732	.002	−0.929 [−1.50, −0.59]
Trustworthiness	4.81 [4.29, 5.34]	4.83 [4.28, 5.39]	0.002	0.281	.961	0.015 [−0.51, 0.54]
Experimenter 3 (<i>N</i> = 60); experimental effect size: Cohen's <i>d</i> = 0.492, 95% CI = [−0.04, 1.00]						
Attractiveness	5.23 [4.84, 5.63]	5.60 [5.20, 6.00]	1.706	0.535	.197	0.347 [0.00, 0.77]
Competence	5.70 [5.30, 6.10]	5.70 [5.30, 6.10]	< 0.001	0.262	1.000	0.000 [−0.37, 0.41]
Friendliness	5.47 [5.07, 5.86]	6.17 [5.77, 6.56]	6.303	3.451	.015	0.659 [0.39, 1.14]
Trustworthiness	5.17 [4.79, 5.54]	5.37 [4.99, 5.74]	0.569	0.333	.454	0.198 [−0.11, 0.62]
Experimenter 4 (<i>N</i> = 50); experimental effect size: Cohen's <i>d</i> = 0.867, 95% CI = [0.25, 1.47]						
Attractiveness	4.19 [3.67, 4.72]	5.17 [4.62, 5.71]	6.662	3.953	.013	0.749 [0.31, 1.33]
Competence	5.81 [5.35, 6.27]	5.88 [5.40, 6.35]	0.042	0.288	.839	0.061 [−0.43, 0.48]
Friendliness	5.58 [5.06, 6.09]	6.04 [5.51, 6.58]	1.591	0.541	.213	0.361 [−0.17, 0.85]
Trustworthiness	5.00 [4.53, 5.47]	5.75 [5.26, 6.24]	4.919	2.024	.031	0.640 [0.21, 1.14]
Experimenter 5 (<i>N</i> = 60); experimental effect size: Cohen's <i>d</i> = 0.759, 95% CI = [0.22, 1.29]						
Attractiveness	4.10 [3.54, 4.66]	4.20 [3.64, 4.76]	0.063	0.269	.802	0.066 [−0.48, 0.62]
Competence	5.87 [5.42, 6.31]	5.90 [5.46, 6.35]	0.011	0.264	.916	0.025 [−0.34, 0.52]
Friendliness	3.27 [2.75, 3.78]	4.10 [3.59, 4.61]	5.273	2.291	.025	0.600 [0.01, 0.10]
Trustworthiness	4.53 [4.08, 4.99]	5.30 [4.85, 5.76]	5.697	2.712	.020	0.628 [0.14, 1.03]
Experimenter 6 (<i>N</i> = 60); experimental effect size: Cohen's <i>d</i> = 0.520, 95% CI = [−0.01, 1.05]						
Attractiveness	4.07 [3.42, 4.72]	4.26 [3.63, 4.88]	0.177	0.283	.676	0.111 [−0.47, 0.78]
Competence	6.28 [5.92, 6.64]	6.03 [5.68, 6.38]	0.942	0.280	.336	−0.262 [−0.64, 0.04]
Friendliness	3.69 [3.17, 4.21]	4.48 [3.98, 4.98]	4.897	1.970	.031	0.577 [0.04, 1.02]
Trustworthiness	4.72 [4.28, 5.17]	5.39 [4.96, 5.82]	4.594	1.744	.036	0.570 [0.12, 0.97]
Experimenter 7 (<i>N</i> = 59); experimental effect size: Cohen's <i>d</i> = 0.212, 95% CI = [−0.31, 0.74]						
Attractiveness	4.47 [3.94, 4.99]	4.55 [4.02, 5.08]	0.052	0.270	.820	0.057 [−0.43, 0.60]
Competence	5.30 [4.83, 5.77]	5.62 [5.14, 6.10]	0.909	0.387	.344	0.252 [−0.19, 0.74]
Friendliness	3.53 [3.00, 4.07]	3.48 [2.94, 4.02]	0.018	0.210	.894	−0.035 [−0.57, 0.49]
Trustworthiness	5.07 [4.66, 5.48]	5.24 [4.83, 5.66]	0.360	0.235	.551	0.155 [−0.16, 0.62]
Experimenter 8 (<i>N</i> = 101); experimental effect size: Cohen's <i>d</i> = 0.517, 95% CI = [0.12, 0.92]						
Attractiveness	3.42 [2.95, 3.89]	3.73 [3.25, 4.20]	0.850	0.306	.359	0.188 [−0.27, 0.65]
Competence	5.48 [5.03, 5.93]	5.47 [5.05, 5.89]	< 0.001	0.210	.975	−0.007 [−0.42, 0.43]
Friendliness	4.96 [4.47, 5.45]	5.92 [5.63, 6.21]	11.463	29.038	.001	−0.675 [−1.15, −0.39]
Trustworthiness	5.40 [4.98, 5.82]	5.90 [5.63, 6.17]	4.076	1.263	.046	0.404 [0.14, 0.82]
Experimenter 9 (<i>N</i> = 99 ^a); experimental effect size: Cohen's <i>d</i> = 0.560, 95% CI = [0.15, 0.97]						
Attractiveness	4.94 [4.47, 5.41]	5.36 [4.96, 5.76]	1.899	0.490	.171	0.279 [−0.11, 0.74]
Competence	6.22 [6.00, 6.45]	6.44 [6.22, 6.66]	1.832	0.476	.179	0.280 [0.06, 0.50]
Friendliness	5.47 [5.08, 5.86]	6.12 [5.85, 6.39]	7.500	5.514	.007	0.464 [0.06, 0.85]
Trustworthiness	5.29 [4.91, 5.67]	6.24 [5.98, 6.50]	17.472	327.991	< .001	0.845 [0.59, 1.22]

(continued)

Table 1. (continued)

Trait	Belief condition		Comparison of belief conditions			
	Low power (<i>M</i>)	High power (<i>M</i>)	<i>F</i>	BF ₁₀	<i>p</i>	Cohen's <i>d</i>
Experimenter 10 (<i>N</i> = 100); experimental effect size: Cohen's <i>d</i> = 0.649, 95% CI = [0.24, 1.06]						
Attractiveness	4.14 [3.83, 4.45]	4.98 [4.67, 5.29]	14.527	101.477	< .001	0.768 [0.46, 1.08]
Competence	6.08 [5.82, 6.34]	6.18 [5.91, 6.45]	0.288	0.240	.593	0.109 [−0.16, 0.36]
Friendliness	5.54 [5.19, 5.89]	6.36 [6.13, 6.59]	15.550	153.281	< .001	0.795 [0.57, 1.14]
Trustworthiness	5.52 [5.16, 5.88]	6.14 [5.89, 6.39]	7.948	6.661	.006	0.570 [0.33, 0.93]
Experimenter 11 (<i>N</i> = 100); experimental effect size: Cohen's <i>d</i> = 0.013, 95% CI = [−0.38, 0.41]						
Attractiveness	3.82 [3.30, 4.34]	3.92 [3.43, 4.41]	0.079	0.218	.779	0.057 [−0.42, 0.57]
Competence	6.22 [5.98, 6.46]	6.10 [5.83, 6.37]	0.446	0.257	.506	−0.135 [−0.40, 0.10]
Friendliness	5.92 [5.58, 6.26]	5.66 [5.20, 6.12]	0.833	0.305	.364	−0.184 [−0.63, 0.15]
Trustworthiness	6.00 [5.71, 6.29]	5.72 [5.30, 6.14]	1.200	0.359	.276	−0.221 [−0.63, 0.07]

Note: Values in brackets are 95% confidence intervals (CIs). Experimenters 1 and 2 participated in Experiment 2; Experimenters 3 and 4 participated in Experiment 3; Experimenters 5, 6, and 7 participated in Experiment 4; and Experimenters 8 through 11 participated in Experiment 5. The effect size (Cohen's *d* and 95% CI) achieved by each experimenter depended on his or her belief about the prime condition. Experimenter effect size is the difference in participants' task performance when the experimenter believed them to be in the high-power versus low-power condition.

General Discussion

In Experiment 1, under double-blind conditions, we failed to find predicted effects of a social power prime on a flanker task, despite robust differences in participants' experiences of power. Results of Experiments 2 through 5 provide consistent evidence that experimenters, rather than prime conditions, influenced target-task outcomes, albeit inadvertently. These results show that subtly revealed expectations can shape other people's behavior and suggest that experimenters are a more powerful stimulus than many researchers, ourselves included, might care to imagine.

Of course, there are a number of possible explanations for why we failed to find priming effects, one of these being task choice. Although in Experiments 4 and 5, we attempted to directly replicate findings in the literature using the same prime and target tasks (Smith & Trope, 2006, Experiment 2, and Smith & Bargh, 2008, Experiment 2), in other experiments, we used variations on reported studies. Whereas in our Experiment 2, we used the same target task as Smith and Trope (2006, Experiment 1), these authors primed power with a writing exercise rather than the scrambled-sentences task. Our Experiment 3 risk-taking measure has not, to our knowledge, been used in power priming, although research has found power-priming effects on similar sequential risk-taking tasks (Jordan, Sivanathan, & Galinsky, 2011; Maner et al., 2007). However, if previously reported power effects are as generalizable as commonly claimed (e.g., Guinote, 2007; Maner et al., 2007; Overbeck & Park, 2006; Smith et al., 2008; Smith & Trope, 2006), the power prime should have influenced

behavior on these tasks. Given that Experiments 1, 3, 4, and 5 showed expected power effects in the manipulation check and that experimenter effects were sensitively detected in Experiments 2 through 5, we do not believe that task choice is responsible for our failure to replicate (conceptually or directly) previous findings.

Another difference between our methods and typical designs is that participants did the manipulation check immediately pre- and postprime. Pilot testing suggested that this was the most reliable way to detect manipulation-related effects. However, it is possible that this procedure contributed to our failure to find a priming effect (e.g., Loersch & Payne, 2012). While additional experimentation is necessary to establish whether priming effects are observed under double-blind conditions without manipulation checks, previous research has found intact priming effects immediately following a manipulation check (e.g., Storbeck & Clore, 2008). Furthermore, power-related test items were hidden within a mood measure, which itself has been shown not to influence priming results when used in this way (Smith & Bargh 2008). Finally, if this manipulation check eliminated the power-priming effect, why did it not also eliminate the experimenter effect?

In contrast, our data suggest that experimenters' expectations about task outcomes influenced participants' performance. This influence was likely exerted via alteration of experimenter behavior, as revealed by experimenter ratings. Although exploratory, these results suggest that effects commonly attributed to priming tasks (e.g., better executive cognition, increased risk taking) might be caused by inadvertent differences in the behavior of non-double-blind experimenters.

Therefore, we believe that this set of findings clearly demonstrates the need for double-blind designs, insofar as this is possible, and explicit measurement of experimenter behavior where it is not.

Note that we do not claim that these results invalidate priming research generally, because they do not show that priming tasks must fail under double-blind conditions. Indeed, reports suggest that priming may work when no experimenter is present (e.g., online; Scholl & Sassenberg, 2015). However, our results do reveal a consistent and unexpectedly powerful influence of experimenter belief communicated during a scripted 5-min interaction. These results suggest that research reports should be regarded skeptically unless authors explicitly report strong double blinding, in which it is impossible for experimenters to become aware of participants' conditions during data collection.

More broadly, our findings suggest that one person's behavior in a social interaction may depend strongly on the beliefs of his or her interaction partner. For example, people's expectations may shape both their own behavior and their responses to others (Snyder & Stukas, 1999). Interaction partners may use behavioral cues to infer another's expectations, thereby allowing themselves to be nudged toward a particular behavior or outcome (Miller & Turnbull, 1986). At a societal level, this result has important implications for understanding how self-fulfilling prophecies arise. For example, teachers may inadvertently favor male students in mathematics and female students in English, leading to gender differences in literacy and numeracy (Nguyen & Ryan, 2008). Thus, these results suggest that understanding the interdependence between social partners' beliefs and behaviors may be important in understanding some intergroup and interpersonal conflicts that arise.

Despite its broad implications, this work has limitations. Because experimenters were exploring priming effects using predictions from the literature, we did not attempt to directly manipulate experimenters' prior beliefs (e.g., inducing experimenters to believe that a high-power prime would impair abstract-thinking ability), although previous research has shown that directly altering experimenter beliefs has a similar effect (Doyen et al., 2012). Additionally, we were unable to explicitly examine the specific behaviors that changed experimenter ratings, because we could not directly observe experimenters without alerting them to the manipulation.

Conclusions

The findings of these experiments have two important implications. First, they suggest that in order to ensure the integrity of research outputs, authors should carefully consider the potential for experimenter effects

during the study-design process and take action to prevent these effects (e.g., video-based participant instruction). Second, these findings suggest that people's beliefs about their interaction partners or about the outcomes of their interactions exert a powerful influence on both interaction-level processes and interaction partners' subsequent behavior. Thus, people's beliefs, stereotypes, and expectations may determine the nature, quality, and outcomes of their interactions.

Action Editor

D. Stephen Lindsay served as action editor for this article.

Author Contributions

T. S. E. Gilder and E. A. Heerey were jointly involved in study conceptualization and data collection and analysis. They jointly wrote the first draft of the article. E. A. Heerey programmed the computer tasks and supervised all students.

Acknowledgments

We thank the students who served as experimenters.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

We thank the Welsh Institute of Cognitive Neuroscience for financial support.

Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797617737128>

Open Practices



All data have been made publicly available via the Open Science Framework and can be accessed at <https://osf.io/pnvjf/>. The design and analysis plans for the experiments were preregistered at <https://osf.io/c4qnz/register/565fb3678c5e4a66b5582f67>. The complete Open Practices Disclosure for this article can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797617737128>. This article has received badges for Open Data and Preregistration. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.

Notes

1. Because we chose to plot individual data points, readers may note the presence of outliers in some figures. Excluding these participants did not substantially change the findings.

2. We explain our approach to ethical issues pertaining to having misled student researchers in the Supplemental Material.
3. See the Supplemental Material for additional notes.

References

- Anderson, C., & Galinsky, A. D. (2006). Power, optimism, and risk-taking. *European Journal of Social Psychology*, 36, 511–536.
- Barber, T. X. (1978). Expecting expectancy effects: Biased data analyses and failure to exclude alternative interpretations in experimenter expectancy research. *Behavioral and Brain Sciences*, 1, 388–390.
- Bargh, J. A., & Chartrand, T. L. (2014). The mind in the middle: A practical guide to priming and automaticity research. In H. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 311–344). New York, NY: Cambridge University Press.
- Doyen, S., Klein, O., Pichon, C. L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PLOS ONE*, 7(1), Article e29081. doi:10.1371/journal.pone.0029081
- Dreisbach, G., & Boettcher, S. (2011). How the social-evaluative context modulates processes of cognitive control. *Psychological Research*, 75, 143–151.
- Eriksen, B., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16, 143–149.
- Fan, E. T., & Gruenfeld, D. H. (1998). When needs outweigh desires: The effects of resource interdependence and reward interdependence on group problem solving. *Basic and Applied Social Psychology*, 20, 45–56.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149–1160.
- Figner, B., Mackinlay, R. J., Wilkening, F., & Weber, E. U. (2009). Affective and deliberative processes in risky choice: Age differences in risk taking in the Columbia Card Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 709–730. doi:10.1037/a0014983
- Galinsky, A. D., Gruenfeld, D. H., & Magee, J. C. (2003). From power to action. *Journal of Personality and Social Psychology*, 85, 453–466. doi:10.1037/0022-3514.85.3.453
- Galinsky, A. D., Magee, J. C., Gruenfeld, D. H., Whitson, J. A., & Liljenquist, K. A. (2008). Power reduces the press of the situation: Implications for creativity, conformity, and dissonance. *Journal of Personality and Social Psychology*, 95, 1450–1466. doi:10.1037/a0012633
- Galinsky, A. D., Magee, J. C., Inesi, M. E., & Gruenfeld, D. H. (2006). Power and perspectives not taken. *Psychological Science*, 17, 1068–1074. doi:10.1111/j.1467-9280.2006.01824.x
- Guinote, A. (2007). Power affects basic cognition: Increased attentional inhibition and flexibility. *Journal of Experimental Social Psychology*, 43, 685–697.
- Harris, C. R., Coburn, N., Rohrer, D., & Pashler, H. (2013). Two failures to replicate high-performance-goal priming effects. *PLOS ONE*, 8(8), Article e72467. doi:10.1371/journal.pone.0072467
- Herr, P. M., Sherman, S. J., & Fazio, R. H. (1983). On the consequences of priming: Assimilation and contrast effects. *Journal of Experimental Social Psychology*, 19, 323–340.
- Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *Journal of Problem Solving*, 7, 2–9. doi:10.7771/1932-6246.1167
- JASP Team. (2017). JASP (Version 0.8.2) [Computer software]. Retrieved from <https://jasp-stats.org/>
- Jordan, J., Sivanathan, N., & Galinsky, A. D. (2011). Something to lose and nothing to gain: The role of stress in the interactive effect of power and stability on risk taking. *Administrative Science Quarterly*, 56, 530–558.
- Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review*, 110, 265–284.
- Klein, O., Doyen, S., Leys, C., Magalhaes de Saldanha da Gama, P. A., Miller, S., Questienne, L., & Cleeremans, A. (2012). Low hopes, high expectations: Expectancy effects and the replicability of behavioral experiments. *Perspectives on Psychological Science*, 7, 572–584. doi:10.1177/1745691612463704
- Loersch, C., & Payne, B. K. (2012). On mental contamination: The role of (mis)attribution in behavior priming. *Social Cognitive and Affective Neuroscience*, 30, 241–252.
- Magee, J. C., Galinsky, A. D., & Gruenfeld, D. H. (2007). Power, propensity to negotiate, and moving first in competitive interactions. *Personality and Social Psychology Bulletin*, 33, 200–212. doi:10.1177/0146167206294413
- Maner, J. K., Gailliot, M. T., Butz, D. A., & Peruche, B. M. (2007). Power, risk, and the status quo: Does having authority promote riskier or more conservative decision-making? *Personality and Social Psychology Bulletin*, 33, 451–462.
- Miller, D. T., & Turnbull, W. (1986). Expectancies and interpersonal processes. *Annual Review of Psychology*, 37, 233–256.
- Nguyen, H. H., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, 93, 1314–1334. doi:10.1037/a0012702
- Overbeck, J. R., & Park, B. (2006). Powerful perceivers, powerless objects: Flexibility of powerholders' social attention. *Organizational Behavior and Human Decision Processes*, 99, 227–243.
- Pashler, H., Coburn, N., & Harris, C. R. (2012). Priming of social distance? Failure to replicate effects on social and food judgments. *PLOS ONE*, 7(8), Article e42510. doi:10.1371/journal.pone.0042510
- Rosenthal, R. (1994). Science and ethics in conducting, analyzing, and reporting psychological research. *Psychological Science*, 5, 127–134.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2001). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.
- Scholl, A., & Sassenberg, K. (2015). Better know when (not) to think twice: How social power impacts prefactual thought. *Personality and Social Psychology Bulletin*, 41, 159–170. doi:10.1177/0146167214559720
- Shanks, D. R., Newell, B. R., Lee, E. H., Balakrishnan, D., Ekelund, L., Cenac, Z., . . . Moore, C. (2013). Priming intelligent behavior: An elusive phenomenon. *PLOS ONE*, 8(4), Article e56515. doi:10.1371/journal.pone.0056515

- Sheldrake, R. (1998). Experimenter effects in scientific research: How widely are they neglected? *Journal of Scientific Exploration*, 12(1), 73–78.
- Sherman, R. A., Rauthmann, J. F., Brown, N. A., Serfass, D. G., & Jones, A. B. (2015). The independent effects of personality and situations on real-time expressions of behavior and emotion. *Journal of Personality and Social Psychology*, 109, 872–888. doi:10.1037/pspp0000036
- Smith, P. K., & Bargh, J. A. (2008). Nonconscious effects of power on basic approach and avoidance tendencies. *Social Cognition*, 26, 1–24.
- Smith, P. K., Jostmann, N. B., Galinsky, A. D., & van Dijk, W. W. (2008). Lacking power impairs executive functions. *Psychological Science*, 19, 441–447. doi:10.1111/j.1467-9280.2008.02107.x
- Smith, P. K., & Trope, Y. (2006). You focus on the forest when you're in charge of the trees: Power priming and abstract information processing. *Journal of Personality and Social Psychology*, 90, 578–596. doi:10.1037/0022-3514.90.4.578
- Snyder, M., & Stukas, A. A., Jr. (1999). Interpersonal processes: The interplay of cognitive, motivational, and behavioral activities in social interaction. *Annual Review of Psychology*, 50, 273–303.
- Storbeck, J., & Clore, G. L. (2008). The affective regulation of cognitive priming. *Emotion*, 8, 208–215.
- Vallacher, R. R., Wegner, D. M., & Somoza, M. P. (1989). That's easy for you to say: Action identification and speech fluency. *Journal of Personality and Social Psychology*, 56, 199–208.
- Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45, 1191–1207.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063–1070.
- Wheeler, S. C., & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin*, 127, 797–826.
- Willis, G. B., Rodríguez-Bailón, R., & Lupiáñez, J. (2011). The boss is paying attention: Power affects the functioning of the attentional networks. *Social Cognition*, 29, 166–181.