The Leadership Quarterly 32 (2021) 101496

Contents lists available at ScienceDirect

The Leadership Quarterly

journal homepage: www.elsevier.com/locate/leaqua

The problem of demand effects in power studies: Moving beyond power priming

Mahshid Khademi *, Marianne Schmid Mast, Christian Zehnder, Oriana De Saint Priest

Faculty of Business and Economics, University of Lausanne, Department of Organizational Behavior, Switzerland

ARTICLE INFO

Article history:
Received 15 May 2020
Received in revised form 25 December 2020
Accepted 31 December 2020
Available online 1 February 2021

Keywords:
Demand effects
Power priming
Resource allocation
Manipulation check
Consequential design

ABSTRACT

Power in experimental research has been commonly induced by methods that raise concerns regarding demand effects. In this paper, we investigate the empirical relevance of these concerns. In an incentivized online study ($N = 1632$), we manipulated the method of power manipulation (power priming vs. resource allocation), the level of power (high-power vs. control), and the presence of a manipulation check after the power manipulation. We then assessed risk-taking as an outcome variable in two ways, once as a non-consequential measure (self-report measure) and twice as a consequential measure (incentivized behavioral choices). Our results show that both using power priming (vs. resource allocation) and implementing a manipulation check substantially increased the potential for demand effects as measured by the proportion of participants who were aware of the study hypothesis. In addition, we were able to replicate the positive effect of power on risk-taking previously reported in the literature. However, we only found a significant (and small) effect for our non-consequential measure of risk-taking; when risk-taking was measured with either of our two consequential measures, power had no significant impact. Our pattern of results shows that concerns about demand effects in priming studies cannot be dismissed. We advise researchers, especially those studying power, to steer away from demand-prone manipulations of power and to measure outcome variables (e.g., behavior) through consequential choices.

Leadership and power are topics that are predominantly studied in separate literatures. Although leadership is often naturally associated with power, little research to date has explicitly addressed this link and its implications (e.g., Bendahan, Zehnder, Pralong, & Antonakis, 2015; Doldor, 2017; Gordon, 2002). One reason for this lack of integration is that research on power and research on leadership are fragmented across different disciplines, which complicates communication. The power literature discusses several consequences of power that are highly relevant for leadership studies. For example, there are studies arguing that people endowed with power tend to ignore advice (de Wit, Scheepers, Ellemers, Sassenberg, & Scholl, 2017; Tost, Gino, & Larrick, 2012), become more creative (Galinsky, Magee, Whitson, & Liljenquist, 2008; Gervais, Guinote, Allen, & Slabu, 2013), are more willing to take risks and to engage in assertive actions (Anderson & Galinsky, 2006; Galinsky, Gruenfeld, & Magee, 2003), and may be prone to engage in immoral behaviors (Bendahan et al., 2015; Giurge, van Dijke, Zheng, & De Cremer, 2020; Lammers, Stapel, & Galinsky, 2010). In light of the importance of these dimensions for people in leadership positions, leadership research can strongly benefit from the power literature.

Moreover, in the power literature, power is often experimentally manipulated in that participants are primed with high-power and compared to a control group or to participants primed with low-power. This setting is used to address what happens if a person obtains or is given power, mirroring the situation of a person who climbs the corporate ladder and is entrusted with more and more leadership tasks. So, potentially, power studies are very informative for the question of what happens to a person when they become a leader.

However, we can only draw valid inferences from power research for leadership if the methods used to investigate power effects are sound and strong. The aim of this paper is to provide evidence for the empirical relevance of these methodological concerns and to propose more robust, alternative methods to do research on power – and thus leadership—in the future.

Currently, experimental investigations that exogenously manipulate power to establish causal effects mainly rely on power priming (Schaerer, Lee, Galinsky, & Thau, 2018). Power priming aims at activating a high or low-power *mindset* in participants. For this purpose, participants are asked to think of either a high- or a low-power role. They either act out this role in a role-play, imagine being that person, or are asked to recall a situation in which they happened to be in the corresponding role. For instance, in typical role-plays, participants are

* Corresponding author.
E-mail addresses: mkhadem@ethz.ch (M. Khademi), Marianne.Schmidmast@unil.ch (M. Schmid Mast), Christian.Zehnder@unil.ch (C. Zehnder), oriana.desaintpriest@unil.ch (O. De Saint Priest).

https://doi.org/10.1016/j.leaqua.2021.101496
1048-9843/

assigned to either the role of a manager (high-power) or the role of a subordinate (low-power; e.g., Anderson & Berdahl, 2002; Briñol, Petty, & Stavraki, 2012). Other manipulations invite participants to imagine themselves in a predefined role (boss or employee) and to write about the feelings and actions related to that role (Dubois, Rucker, & Galinsky, 2010; Schmid, 2018).

The probably most widely used priming technique is the so-called recall task in which participants are asked to write about a situation in their past in which they had power over others (for high-power priming) or in which someone else had power over them (for low-power priming). In the power-neutral condition, participants write about an event that happened to them the day before (Galinsky et al., 2003). The idea behind this manipulation is that recollection of memories activates a power-related mindset that then affects behavior and decision-making.

Despite their popularity, using such power priming methods[1] might be problematic in that they potentially create *demand effects* (Sturm & Antonakis, 2015; Schaerer et al., 2018). Demand effects refer to experimenter-induced cues and expectations in the context of research experiments which may influence the behavior of participants (Orne, 1962). Such cues can, for example, be embedded in the description of the experiment (if instructions to participants reveal the research question) (Charness, Gneezy, & Kuhn, 2012; Zizzo, 2010) or may emerge if a manipulation check is performed before the dependent variable is measured (Chester & Lasko, 2019; Hauser, Ellsworth, & Gonzalez, 2018; Lonati, Quiroga, Zehnder, & Antonakis, 2018).

Power priming methods require the explicit mentioning of the manipulated variable (power) in the treatment conditions. Such manipulations may therefore reveal critical information about the hypothesis of the study and may induce participants to adapt their behavior to what they infer being the researcher's expectations. As a consequence, demand-prone methods may yield confounded results; it remains unclear to what extent the ultimate responses are the result of what powerful or powerless people *actually* do or if they reflect what people think powerful or powerless people are *expected* to do (i.e., demand effects).

Another drawback of existing power priming methods is that, most of the time, being in a powerful or powerless psychological state does not entail experiencing actual power (Flynn, Gruenfeld, Molm, & Polzer, 2011; Smith & Hofmann, 2016). For instance, if the powerholder is in charge of evaluating the performance of the powerless or selecting a specific task for them, these choices typically neither affect outcomes for the powerholder nor for the followers. When reflecting on the current situation in the power literature Smith and Hofmann (2016, p. 10043) wrote:"*Although such manipulations allow for causal attributions, the most common ones involve thinking about power or anticipating power differences rather than experiencing them. Even when participants experience low- and high-power roles in the laboratory, these roles generally do not involve real decisions or consequential outcomes*." Power priming does not manipulate "actual power" but rather "felt power". That is, the operationalization of power in power priming does not entail actual control over valued resources and thus is not aligned with the prevalent definition of power. Indirect approaches of manipulating power such as recalling a power-related memory or playing a role in a hypothetical scenario are problematic, because it is not clear whether actual power and felt power affect behavior in the same way. Moreover, it is important that researchers are aware that power has many possible sources (e.g., control over resources, knowledge (Raven, 1993) and that different sources may lead to different forms of power with (potentially) different implications. This variety in types of power implies that power manipulations need to be adapted to the forms and sources of power that are relevant for a particular research question. Of course, it is possible that for certain questions, the concept of interest may indeed be "felt" power. For example, Goldstein and Hays (2011) investigate the effects of what they call "illusory power transference" (feeling powerful because of an association with a powerful other). In such cases, it is fully legitimate to manipulate felt power. However, if the relevant concept is a form of actual power, a convincing manipulation needs to directly target this variable.

Indirectly manipulating power also creates a second disadvantage: the lack of an objective change in actual power requires testing whether the manipulation subjectively affected felt power in participants (participants in the high-power condition should report feeling more powerful than those in the low-power or neutral condition). In many studies using power priming (Galinsky et al., 2003; Guinote, 2007a; Guinote, 2007b; Smith, Jostmann, Galinsky, & van Dijk, 2008), the manipulation check is executed right after the power manipulation and before the dependent variable is measured. Such a procedure reinforces the potential for demand effects, because the manipulation check may provide strong hints on the manipulated variable.

Finally, there is an additional reason—unrelated to the use of power priming technique - for why power studies may be particularly prone to demand effects. Power research as a sub-domain of social psychology is one of the fields that often uses self-report measures or hypothetical scenarios to measure outcomes of interest (e.g., participants are asked to describe what they would do in a particular situation) (Sassenberg & Ditrich, 2019; Schaerer et al., 2018). Such non-consequential operationalizations of the dependent variable (self-report measures or hypothetical scenarios to measure outcomes of interest) are especially likely to be affected by what people think is expected of them in a given scenario, because deviations from true behavioral intentions have no costs for participants (Antonakis, 2017; Lonati et al., 2018).

In a nutshell, with regard to demand effects, we argue that using power priming and manipulation checks of this priming can create a large potential for demand effects. Moreover, non-consequential outcome measures make it likely that the demand-effect potential has an effect on results, because participants possibly respond according to demand. To the best of our knowledge, there is not yet any empirical work that explicitly measures the extent to which power priming and the corresponding manipulation checks affect the potential for demand effects in power studies and how more or less consequential measures of outcomes are affected by different power manipulations. In this paper, we make a first attempt at filling this gap in the literature.

For this purpose, we compare an established and often used power priming method with an alternative power manipulation that directly affects actual power in that it provides powerholders with *asymmetric control over valued resources*, reflecting the definition widely used in research (e.g., Emerson, 1962; French & Raven, 1959; Goldstein & Hays, 2011; Gruenfeld, Inesi, Magee, & Galinsky, 2008; Jordan, Sivanathan, & Galinsky, 2011; Keltner, Gruenfeld, & Anderson, 2003; Magee & Galinsky, 2008; Rucker, Dubois, & Galinsky, 2011). Giving actual control over others' resources (resource allocation) as the power manipulation has two immediate advantages. First, we do not need to explicitly mention the manipulated variable. Second, because there is an objective difference in the degree of participant's control (i.e., an objective difference in actual power), our manipulation does not necessitate a subsequent manipulation check.

Moreover, we disentangle the effects of the power manipulation and the manipulation check using a 2 by 2 design in which we cross two different power manipulations (power priming and resource allocation) with the presence/absence of a manipulation check. In a first step, we investigate the potential for demand effects under each of the four conditions. To this end, we measure how many participants correctly guessed the hypothesis of the study (potential for demand effects). In a second step, we investigate the impact of the type of power manipulation and the presence/absence of a manipulation check on the link between power and a dependent variable that has been repeatedly studied in the power literature: risk-taking. In order to gain insight into the difference between non-consequential and consequential

[1] We focus on explicit and not on implicit power priming methods because the former are the ones used more widely and the ones most prone to demand effects. When we write "power priming" we mean "explicit" priming of power.

outcome variables, we use versions of both consequential and non-consequential measures of risk-taking.

Power priming and potential for demand effects

As mentioned before, power is typically defined as asymmetric control over valued resources (Magee & Galinsky, 2008). However, prevalent methods of power priming often give no actual control to powerholders and by manipulating felt instead of actual power, the chances of generating demand effects are increased (Antonakis, 2017; Sturm & Antonakis, 2015). This is not the first time that power priming methods are under scrutiny. Certain power priming methods have been shown to produce spurious results. A well-known example is power posing (Carney, Cuddy, & Yap, 2010). Replication studies found no effect of striking a power pose on either cortisol or risk-taking, or any other dependent variable, except for self-reported felt power (Jonas et al., 2017; Ranehill et al., 2015). Moreover, doubts concerning the robustness of priming procedures in general have been voiced by many scholars (Chivers, 2019; Kahneman, 2012) as a result of failed attempts at replicating well-established findings in the social priming literature (e.g., Doyen, Klein, Pichon, & Cleeremans, 2012; Gilder & Heerey, 2018; Harris, Coburn, Rohrer, & Pashler, 2013; Pashler, Rohrer, & Harris, 2013; Shanks et al., 2013).

In power priming (e.g., role-play of high- and low-power roles, imagine being a high- or low-power person, or think about a situation in the past in which a person had power or not), the goal is to put the person in a high- or low-power mindset; in other words having them *feel* powerful or powerless (or neutral with respect to power). The effect of the priming technique is rather unconscious and works through a mental representation of the primed concept (Bargh & Chartrand, 1999). In other words, the prime activates a cognitive network in the powerholder and as a consequence, several learnt associations in relation to the prime become available and affect an individual's feelings and behavior. For example, in power priming, the recalled memory about a power incident or the role-play activates a cognitive network of power that brings about the emotional, cognitive, and behavioral consequences of being in power, despite the person primed not being granted actual power or control over valued resources (Tost, 2015).

The majority of empirical studies in the power field rely at least to some extent on experiential power manipulations that include role-playing, episodic recall tasks, or asking participants to take part in imaginary role-plays (Schaerer, du Plessis, Yap, & Thau, 2018; Schaerer et al., 2018; Sturm & Antonakis, 2015; Tost, 2015). A review of 399 experiments investigating power consequences up until 2015 showed that more than half (54%) of the manipulations used power priming. Power priming methods require that the experimenter mentions a power role or a power-related incident in the instructions. Participants are therefore confronted with salient cues about the manipulated variable. To the extent that participants have expectations about how a powerful or a powerless person ought to behave in a particular situation, these beliefs might guide how participants will react, behave, or decide in the subsequent task. This mechanism can be problematic because the observed effect of power on the behavior of interest is not driven by a change in actual power, but may simply reflect what people think the effect should be (demand effect). We do not know whether people who are really given power would react in the same way.

One of the widely used power priming techniques and "… *by far the most common approach to manipulating power in social psychology over the last decade*…" (Tost, 2015, p. 50) is recall priming. The recall prime is based on the assumption that everyone has relevant memories of power incidents and knows what power entails from personal experience (Galinsky et al., 2003). However, because the researcher has no control over these memories the manipulation is necessarily very subjective. Different participants may recall very different forms of power (e.g., resource power, expert power, referent power etc.). We cannot be sure that these different forms and expectations of power lead to comparable emotional and behavioral responses and the resulting heterogeneity may complicate the interpretation of results considerably (e.g., Hu, Rucker, & Galinsky, 2016; Tost & Johnson, 2019). Some might believe that researchers have more control when using versions of power priming in which participants are asked to recall an incident in a specific leader/follower situation (e.g., Dubois et al., 2010; Rucker et al., 2011; Schmid, 2018). However, such attempts may encounter the problem that at least some participants may not have any relevant prior experience. We therefore argue that for power research to be meaningful for the leadership literature, we need to understand how people who are given actual power will behave. Thus, we used an alternative power manipulation method, called the resource allocation (RA) manipulation. In this method, actual power is given to a person through granting control over valued resources in a structural setting. The manipulation grants the same level of actual control over valued resources to individuals and does not require any subjective previous experience. Individuals in positions of power are free to decide about valued resources for others. This freedom of choice and the control over the resource allocation to others are regarded as dimensions of power (Keltner et al., 2003; Magee & Smith, 2013). The fact of commanding those resources typically is paralleled by a sense of power, or, in other words, by feeling powerful because such power would be simultaneously accompanied with the understanding that other people depend on the powerholder (Tost, 2015). Thus, feeling powerful will be the subjective, emotional part of being given actual power.

Concretely, participants in our RA manipulation find themselves in groups and are randomly assigned the role of an allocator, an observer, or a receiver. Each group has one allocator (high-power role), one observer (neutral role), and five receivers (low-power role). The allocator distributes a number of lottery tickets among the receivers and has absolute control over the distribution of the tickets (power) but cannot take any tickets for themselves. The observer is not involved in the lottery ticket distribution and only observes the distribution of the allocator. The observer has no power to disagree or change the distribution that the allocator makes and has no gain or loss from the distribution. Both the allocator and the observer receive the same fixed payment for participation. The receivers also are compensated with a similar amount, but their expected payment also depends on how many tickets they receive.

The resource allocation method has several advantages. First, we use actual power as in the definition of power (asymmetric control over valued resources). Power is induced using a real endowment (i.e., no deception is used) which not only raises the ecological validity of the design, but also increases the internal validity of the manipulation (correspondence between definition and operationalization of power). This design reduces demand effects because participants do not have to think about what might be expected from them in their power role. They simply act within the role. Also, there is nothing in the instructions that mentions power or any related concept and participants are thus not made aware of the hypothesis of the study. It is important to use neutral wordings because masking the dependent and independent variable lowers the chance of generating demand effects (de Quidt, Haushofer, & Roth, 2018).

Note that demand effects cannot be eliminated entirely from an experimental design (Orne, 1962, 1969). Yet researchers are obliged to minimize the cues that convey the experimental hypothesis to their participants (Rosnow & Rosenthal, 1997), because participants tend to behave in accordance to what they believe the hypothesis of the study is, meaning how the independent variable affects or is related to the dependent variable (Nichols & Maner, 2008). If participant's beliefs match the hypothesis, the threat of demand effects is heightened (Zizzo, 2010); a positive correlation between the correct guesses of the hypotheses and the ultimate behavioral outcome is problematic because it will obscure the interpretation of the results (it is not clear whether the results are caused by the treatment or the demand). In order to understand and measure such potential for demand effects, we followed the traditional "post-experimental inquiries" by simply asking participants what they thought the purpose of the study was (Orne, 1962).

In the current research, we compare our RA manipulation with one of the most widely used power priming methods (Galinsky, Rucker, & Magee, 2015; Gruenfeld et al., 2008; Guinote, 2007a; Lammers, Galinsky, Gordijn, & Otten, 2012; Schmid, Kleiman, & Amodio, 2015; Smith & Trope, 2006; Tost, 2015; Tost, Gino, & Larrick, 2013) in which participants are asked to write an essay about a time when they had power over somebody (high-power condition) or to write about a situation when someone else had power over them (low-power condition), or to write about what they did the day before (neutral condition) (Galinsky et al., 2003).

Hypothesis 1. *More participants correctly guess the research hypothesis of the study (*i.e.*, there is a larger potential for demand effects) under power priming (PP) than under resource allocation (RA).*

Manipulation checks and potential for demand effects

Manipulation checks aim at ensuring that the manipulation was successful (Kidd, 1976; Perdue & Summers, 1986). Manipulation checks contribute to the construct validity of the study (Sigall & Mills, 1998) which refers to the extent to which the intervention operationalizes what it claims to operationalize (Cook, Campbell, & Shadish, 2002). In a survey, 75% of the scholars in social psychology think that "manipulation checks are necessary" in a well-designed study (Fayant, Sigall, Lemonnier, Retsin, & Alexopoulos, 2017, p. 127).

When actual power is manipulated (e.g., through the RA power manipulation), the need for a manipulation check will logically be removed (Lonati et al., 2018), because the objective changes in the elements of the design do not require a subjective interpretation from the participants. However, when using power priming, the activation of the power mindset can vary from one participant to the other, so that scholars are required to conduct a manipulation check to ensure that the relative subjectivity of the mindset priming goes in the desired direction. Often, the check is performed right after the manipulation; this is commonly done by explicitly asking participants how "powerful", "in charge", or "in control" they felt after the essay writing task (Guinote, 2007a; Guinote, 2007b; Rucker et al., 2011; Smith et al., 2008).

When the manipulation check is implemented after the manipulation and before measuring the dependent variable, key information about the study may become salient to the participants (Kühnen, 2010) and the study hypothesis may become transparent (Ejelöv & Luke, 2020; Lonati et al., 2018). Participants might have converging ideas about how a person in power ought to behave, react, and respond, thus introducing demand effects. In other words, sometimes "what we call a manipulation check could also be a manipulation" (Hauser et al., 2018, p. 7). Participants' responses to the manipulation check might affect the outcome variable, because the manipulation check impacts what participants think to be the expected behavior. Such reactivity to a manipulation check can inflate (or deflate) the true causal link between the manipulation and the outcome variable (Ejelöv & Luke, 2020).

Hypothesis 2. *More participants correctly guess the research hypothesis of the study (*i.e.*, there is a larger potential for demand effects) if a manipulation check is performed before measuring the dependent variable.*

Demand effects in the context of power and risk-taking

In our study, we selected "risk-taking" as our dependent variable of interest, because previous research has found a strong relation between power and risk-taking (Anderson & Galinsky, 2006; Galinsky et al., 2003; Jordan et al., 2011; Lewellyn & Muller-Kahle, 2012; Maner, Gailliot, Butz, & Peruche, 2007; Ronay & Von Hippel, 2010). We used one non-consequential and two consequential assessments of risk-taking as outcome measures. When choosing the risk-taking tasks, we pursued two goals. First, we wanted to use what is commonly studied in the literature and second, we wanted to use measures that we thought might be differentially susceptible to demand effects. This is why we chose a non-incentivized self-reported risk-taking task (which we expect to be strongly affected by demand), and two incentivized tasks. As the two consequential tasks we used the Balloon Analog Risk Task (BART) (Lejuez, Aklin, Zvolensky, & Pedulla, 2003; Lejuez et al., 2002) and the lottery task (Dohmen et al., 2011). The BART task is often used in psychology and in neuroscience (Jordan et al., 2011; Lewellyn & Muller-Kahle, 2012) whereas the lottery task is rather used in economics (Forsythe, Horowitz, Savin, & Sefton, 1994; Holt & Laury, 2002). In both of the consequential measures of risk-taking, participants are compensated based on their choices in risky settings. This broad range of tasks should allow us to detect a link between power and risk-taking if there is one because the tasks are different and are likely to tap into different aspects of risk-taking.

Hypothesis 3a. *Power increases the propensity to take risks. Participants in high-power conditions therefore show a higher degree of risk-taking than participants in the control conditions.*

Hypothesis 3a is a pure replication hypothesis and relates to a general comparison between power and control conditions. We will therefore test this hypothesis by pooling our data from all the power manipulation methods (i.e., PP and RA, with/without manipulation check) to test this hypothesis for each of our three measures of risk-taking separately. However, when the treatment entails indications or cues (explicitly or implicitly) about the appropriate/expected behavior, it can ultimately lead participants to behave in accordance with experimental demand - what the participants think the aim of the study is or what they think is the expected behavior on their part (Antonakis, 2017). We now propose that in the conditions where participants have higher chances of guessing the aim of the study correctly (using the PP method and/or explicit manipulation checks) they may adapt their behavior in accordance with the hypothesis of the study. We thus expect that the positive relation between power and risk-taking is particularly strong when power is manipulated in a way that is prone to demand effects.

Hypothesis 3b. *Demand-prone power manipulations lead to a stronger positive impact of power on risk-taking. Specifically, we hypothesize that (i) using power priming (PP) leads to a stronger impact of power on risk-taking as compared to manipulating power through resource allocation (RA) and (ii) performing a manipulation check before measuring the dependent variable will strengthen the impact of power on risk-taking.*

As with Hypothesis 3a, we will also test Hypothesis 3b for each of our three measures of risk-taking separately. However, although we expect Hypothesis 3b to be true for each of our measures of risk-taking, we also predict that the effects of our different power manipulations will diverge particularly when we assess risk-taking with our non-consequential measure.

Whereas non-consequential self-report measures are an appropriate tool to assess individuals' emotions, perceptions, and beliefs about a phenomenon (Podsakoff & Organ, 1986), using non-consequential or hypothetical measures to measure behavior is often problematic. With non-consequential measures, taking action to please the experimenter is not costly (Podsakoff, MacKenzie, & Podsakoff, 2012), entailing considerable room for demand effects to take place (e.g., Durgin et al., 2009). In addition to their susceptibility to demand driven outcomes, non-consequential designs are generally weak in experimental realism (Colquitt, 2008; Podsakoff & Podsakoff, 2019). In consequential settings, in contrast, it is costly for participants not to pick their preferred choice in order to comply with the demand (de Quidt, Vesterlund, & Wilson, 2019; Lonati et al., 2018). In this sense, the use of non-consequential outcome variables makes it more likely that demand effects only stem from a confound of the actually observed effect. We therefore expect to see particularly strong support for Hypothesis 3b when performing the analysis of non-consequentially measured risk-taking.

Method

Pre-tests

Although we detailed that when using a manipulation of actual power instead of a power priming procedure, there is no need for checking whether subjectively, power is experienced differently, we performed a pre-test to measure the impact of both power manipulations, PP and RA, on participant's felt power. The reason for doing so is related to our setting: we need both manipulations to be comparable in strength so that if we find effects, they cannot be interpreted as stemming from a difference in the strength of manipulating felt power. In other words, if we find differences between the effects of the two power manipulations on risk-taking and one manipulation is substantially stronger with respect to felt power, we do not know whether this difference is driven by the manipulation type—which is what we are interested in—or simply by the strength of the manipulation. We therefore needed to ensure that the PP and the RA methods resulted in similar effect sizes on felt power.

Phase one

The goal of phase one of the pretest was to investigate how different levels (high, neutral, low) of each manipulation affect reports of felt power by participants. We used a 2 (power manipulation method: PP vs. RA) by 3 (power level: high-power, neutral, and low-power) design with random assignment to one of the 6 experimental conditions. We recruited 323 participants from MTurk (https://www.mturk.com) and excluded 31 observations due to attention check failure,[2] resulting in a sample of 292 individuals, 47.95% female; $M_{age} = 33.48$, $SD_{age} = 9.16$.

Power priming (PP). Following Galinsky et al. (2003, p. 458), participants in the high-power condition were asked to recall an incident in which they had power over another individual or individuals and write about it. In the low-power condition, participants were asked to write about an incident in which someone else had power over them and in the neutral condition, participants were invited to write about the last time they went to the grocery store.

Resource allocation manipulation (RA). Participants were assigned to groups of seven individuals. There were three roles in each group: allocator (high-power), observer (neutral), and receiver (low-power). There were 1 allocator, 1 observer, and 5 receivers in each group. The roles were distributed randomly among participants. The allocator was given five lottery tickets to distribute among the receivers. The allocator had full control over how to distribute the 5 tickets among the receivers. The only restrictions were that the allocator was not allowed to keep any lottery tickets for him/herself and that all the tickets needed to be distributed. Before the distribution was made, the allocator was informed about the gender and age of the receivers (e.g., receiver A: female, age 25–30; receiver B: male, age 20–25 etc.). The participant in the role of the observer (neutral power) was not involved in the distribution phase and only received the information about the allocator's distribution. The observer could not take any action to change the allocator's decision and had no gain or loss from the distribution. Both the allocator and observer received the same amount of money ($3) for their participation in the study. The receivers also gained the same exact fixed payment ($3). In addition, receivers were paid based on the lottery ticket distribution. A receiver with one lottery ticket had a 1 in 10 chance to win 1 additional dollar. A random device selected 25 out of the 250 distributed lottery tickets and the final gains were paid to participants accordingly. The participants in the different roles were not informed about the final payments of participants in the other roles to avoid salient comparisons. On Qualtrics, we ran the high-power resource allocation condition first (in which allocators distribute lottery tickets) and then with a short time lag we ran the low-power condition (conveying information about the distribution of lottery tickets to the observers). This way we always first stored the decisions of the allocators and then communicated this information to the observers. In a last step, we recruited the receivers and distributed the lottery tickets as was decided by the allocators.

Manipulation check: Measuring felt power. After the power manipulation, participants answered one question about how powerful they felt using a Likert scale from 1 (not at all) to 7 (very much) ($M = 3.77, SD = 1.94$). In the PP condition, high-power participants felt more powerful ($M = 5.16, SD = 1.68$) than participants in the neutral condition ($M = 3.94, SD = 1.59$), $F(1,286) = 13.54$, $p < 0.001$, Cohen's d = 0.75, and than participants in the low-power condition ($M = 2.56, SD = 1.49$), $F(1,286) = 65.73$, $p < 0.001$, Cohen's d = 1.63. In the RA condition, high-power participants felt more powerful ($M = 5.38, SD = 1.27$) than participants in the neutral condition ($M = 2.71, SD = 1.79$), $F(1,286) = 71.63$, $p < 0.001$, Cohen's d = 1.71, and than participants in the low-power condition ($M = 2.88, SD = 1.60$), $F(1,286) = 73.66$, $p < 0.001$, Cohen's d = 1.73 (see Fig. 1).[3]

To assure that the two types of power manipulation have comparable strength, we selected for the PP the high-power versus low-power condition (d = 1.63) and for the RA the high-power versus neutral condition (d = 1.71) for further pretesting and for the main study.[4] There were two reasons why we selected the neutral condition over the low-power condition for the RA method: 1) to have a relatively similar effect size compared to the PP manipulation method and thus to avoid the possibility that the RA manipulation leads to a weaker effect on risk-taking because of a smaller effect size; 2) to avoid income effects between the conditions chosen in the RA method. The allocators and observers obtained the exact same payment by design, whereas the receivers' income depends partially on the number of lottery tickets received from the allocators and partially on chance.

In sum, to conduct the next phase of the pre-test and the main experiment, for the PP condition we kept the high and low-power conditions and for the RA method, we used the high-power and the neutral condition. Note that from now on, we simply call the group to which the high-power individuals are compared, the *control group* because being in the control group elicits a comparable amount of feeling relatively less powerful than being in the high-power group for both types of power manipulations.

[2] There were three attention checks. Failure in at least two of them led to exclusion from the whole study. The items of the attention checks were: "I sleep less than an hour per night", "I eat a computer everyday", and "All my friends are aliens". This type of attention checks are categorized as logical statements addressed by Abbey and Meloy (2017).

[3] In the RA condition, the low-power participants (receivers) felt rather more powerful ($M = 2.88, SD = 1.60$) than the participants in the neutral condition (observers) ($M = 2.71, SD = 1.79$). Although the difference is not significant, two possible reasons come to mind explaining this unexpected pattern of results for the observers: 1) being excluded from the distribution game and not receiving any tickets from the allocator and 2) being faced with the allocator's distribution of the lottery tickets with which the observers did not necessarily agree might have lowered their felt power.

[4] Note that comparing the two low-power conditions in RA ($M = 2.88, SD = 1.60$), and PP ($M = 2.56, SD = 1.49$), gave no significant difference $t(96) = 1.02$, $p = 0.16$, but there was a significant difference between the neutral conditions when comparing neutral power in RA ($M = 2.71, SD = 1.79$), to neutral power in PP ($M = 3.94, SD = 1.59$), $t(94) = 3.53$, $p < 0.001$. But, when we compared the neutral conditions in RA ($M = 2.71, SD = 1.79$), and low-power condition in PP ($M = 2.56, SD = 1.49$) again there was no significant difference $t(95) = 0.45$, $p = 0.33$). That was another reason why we selected the neutral condition from RA in comparison to low-power condition from PP as there were no difference in terms of felt power.

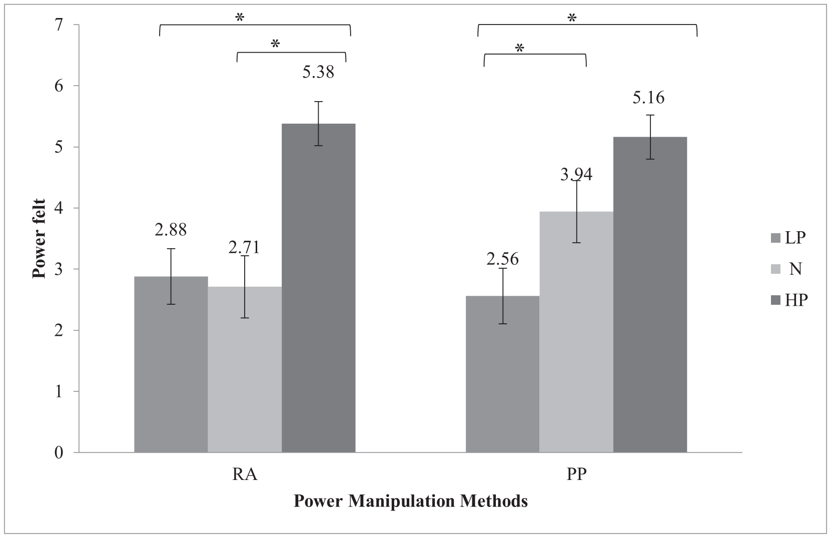

Fig. 1. The effect of power manipulation method on felt power in the pre-test, phase one ($N = 323$). PP = power priming, RA = resource allocation; three power levels: HP = high-power, N = neutral, and LP = low-power. (*) indicates statistically significant difference ($p < 0.01$).

Phase two

In this second phase, our main objective was to confirm that the two manipulations selected in phase one indeed have comparable impacts on felt power using a larger sample of participants. In addition, we also measured how transparent each of the two manipulations was to participants. The transparency of the manipulation is an important pre-requisite for the emergence of a potential for demand effects. If the participants understand what variable the study manipulates, they may be able to correctly guess the hypothesis that the researcher investigates.

On the same online platform, 746 participants were recruited. We excluded 140 participants due to lack of correct responses to the attention checks (same as in phase one). The final sample consisted of 606 participants, 41.42% female; $M_{\text{age}} = 36.13$, $SD_{\text{age}} = 10.86$.

We randomly assigned participants to one of the experimental conditions in a 2 (method of power manipulation: PP vs. RA) by 2 (level of power manipulation: high-power vs. control) between-subjects design. We used the same item as in phase one to measure for how powerful they felt. In the PP condition, high-power participants felt more powerful ($M = 5.08$, $SD = 1.54$) than participants in the control condition ($M = 2.97$, $SD = 1.73$), $F(1,602) = 125.21$, $p < 0.001$, Cohen's d = 1.29. In the RA condition, high-power participants (allocators) felt more powerful ($M = 5.24$, $SD = 1.38$) than participants in the control condition (observers) ($M = 2.56$, $SD = 1.69$), $F(1,602) = 228.10$, $p < 0.001$, Cohen's d = 1.73 (Fig. 2). Observed effect sizes were similar to those identified in phase one. The effect size in the RA condition was slightly larger than in the PP condition. However, if anything, having a stronger RA manipulation works against our predictions, because we hypothesized that the RA power manipulation would have a *smaller* effect on risk-taking than the PP manipulation, (see Hypothesis 3).

To measure the transparency of the manipulation for participants we asked whether they could guess what the study was about in a free-form question. We coded their answers as 1 when it mentioned power or control (e.g., being in control, being in charge, having any saying in a situation) and 0 otherwise. Participants in the PP correctly guessed the manipulated variable more often ($M = 0.44$, $SD = 0.50$) compared to participants in the RA condition ($M = 0.15$, $SD = 0.36$), $t(604) = 8.09$, $p < 0.001$. The fact that the manipulation is more transparent does not necessarily imply a larger potential for demand effects because participants may still fail to correctly guess the research hypothesis. However, a transparent manipulation is a necessary prerequisite for the potential presence of demand effects.

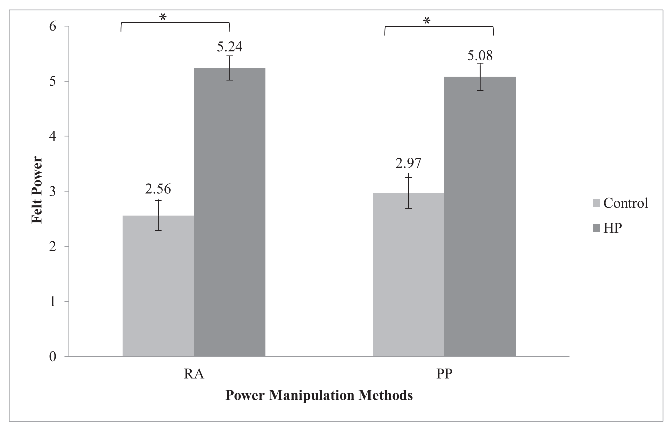

Fig. 2. The effect of power manipulation method on felt power in the pre-test, phase two ($N = 606$). PP = power priming, RA = resource allocation; two power levels: HP = high-power and control. (*) indicates statistically significant difference ($p < 0.01$).

Main experiment

Participants

We recruited 1752 individuals from the same online platform (MTurk). We used the same attention check questions and exclusion criteria. Additionally, we checked the content of the stories written in the PP condition and excluded stories that were either 1) non-relevant to power situations, 2) included pieces of text copy-pasted from the internet, or 3) nonsense responses (Rinderknecht, 2019). This resulted in excluding 119 observations with a final sample consisting of 1633 individuals (43.11% female; $M_{\text{age}} = 35.80$, $SD_{\text{age}} = 10.75$). Participants received a fixed amount of $3. They could gain an additional $10 based on some random elements in the protocol of the risk-taking tasks, described in more detail below.[5]

[5] The fixed payments were paid instantly at the end of the study and participants were informed that the varied payments were paid with a time lag after the study as bonus payments on the Mturk platform. This is because the final payoffs had to be calculated based on the decisions participants made in each step.

Procedure and measures

Participants were randomly assigned to one condition of a 2 (power manipulation method: PP vs. RA) by 2 (level of power: high-power vs. control) by 2 (manipulation check: with vs. without) experimental design (See Table 1). In the first step, we manipulated power using either the PP or the RA power manipulation method described above. We then randomly assigned participants to either the condition with or without manipulation check and participants in the "with manipulation check" condition reported how powerful they felt. We then assessed participant's level of risk-taking with three different measures: i) incentivized BART task, ii) non-consequential self-report, and iii) incentivized lottery task (all measures are explained in detail below). Participants always faced the same sequence of risk-taking tasks. Because we are interested in differences across conditions and because in all conditions, the order was the same, the differences among conditions cannot be explained by order effects. Finally, we measured the potential for demand effects by asking participants to guess the hypothesis of the study.

Power manipulation methods

We used the PP and the RA methods manipulating two levels of power: high-power and control. It is important to recall the main differences between the two methods. PP does not change actual power, but aims at affecting participant's power mindset (i.e., felt power) by inducing them to remember and describe a situation in which they found themselves in a position of power or in a powerless state (depending on the condition). RA directly manipulates actual power by changing participant's impact on the incomes of others. Participants in the high-power condition can choose how many lottery tickets other participants get and participants in the control condition observe the allocation choices of the high-power people.

Manipulation check

We used the presence of a manipulation check as a treatment variation, because we hypothesized that the presence of a manipulation check would increase the potential for demand effects (in other words: we expected participants to be more likely to guess the hypothesis correctly if they had confronted a manipulation check after the actual manipulation). Those who were assigned to the "with manipulation check" condition were asked how powerful they felt with the following question: "After doing this task, please indicate how powerful you feel now on a 7-point Likert scale ranging from 1 (not at all) to 7 (very much)? "In the condition without manipulation check, the participants simply advanced to the next steps of the study.

Self-reported risk-taking

As the non-consequential measurement of risk-taking, we used a self-report measure. Self-reported risk-taking has shown to be a good predictor of risk-taking tendencies in various domains such as job choices and portfolio selections (Dohmen et al., 2011). We asked participants "How do you see yourself: are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?" (Pannenberg, 2010). They had to rate their risk-taking from 1 (unwilling to take risks) to 7 (fully prepared to take risks). Higher scores indicate higher risk-taking ($M = 4.36$; $SD = 1.63$).

Table 1
Number of observations in each condition and each role.

Condition	Without manipulation check	With manipulation check
PP_control	202	192
PP_HP	194	196
RA_control (observer)	212	211
RA_HP (allocator)	209	217

Note. PP = power priming and RA = resource allocation.

BART risk-taking

We also measured risk-taking behavior in a consequential way, using a behavioral measure developed by Lejuez et al. (2002). In the BART task, participants are faced with several rounds of risky decision-making. Participants had to "inflate" 10 (digital) balloons. More specifically, they saw a balloon that they needed to inflate with a pump. Each air blast that they added improved their score by 1 point, but it also increased the risk that the balloon would burst. Participants could inflate the balloon as much as they liked and could collect the points at any time. However, when the balloon burst, all the points accumulated in that specific round were lost. Bursting the balloon did not have an effect on the collected points in former rounds. There were 10 rounds of balloon inflating and the threshold for bursting was different for each balloon. The final payoff was based on the overall collected points (100 points = $1) and participants could gain up to $3.68 in this task. Past research has shown that the points collected via this measure have a significant and reliable relation with risk-taking in various domains including minor crimes, substance use, risky sexual behavior, and unhealthy behavior (Lejuez et al., 2007). People who inflate balloons more, show more risk-taking behavior. We used the average number of air blasts added across all rounds as the BART risk-taking behavior measure ($M = 14.76$; $SD = 7.29$).

Lottery risk-taking

As another consequential way of measuring risk-taking, we used a paid lottery task (Dohmen et al., 2011). In this task, participants were presented with a table of 20 rows. In each row participants could either choose a safe option (a fixed amount of money) or play a lottery (win $0 or $7 with 0.5 probability). The lottery remained unchanged in all rows (both the probability and the amounts stayed the same) but the safe option increased by $0.25 from row to row. So, for example, in the first row the participant had to select between a safe option of $0 and a lottery in which they could win $0 or $7 with 0.5 probability. In the second row, the participants had to choose between a safe option of $0.25 and the lottery (winning $0 or $7 with 0.5 probability). This continued up to the last (twentieth) row, where the participant had to select between a safe option of $4.75 and the lottery (winning $0 or $7 with 0.5 probability). In the first rows the safe option is very unattractive relative to the lottery. Accordingly, most participants decided to play the lottery. At some point, however, when the safe option increases and approaches the expected value of the lottery ($3.5), participants start to switch to the safe option. We use this switching point (i.e., the lowest amount at which participants prefer the safe option) as the measure of risk-taking in this task.

Participants were informed that for 1 out of 7 participants, one row would be randomly selected and the participant's decision in the selected row would be implemented and paid out. Participants were therefore able to gain up to $7. Some participants exhibited choice patterns with multiple switching points. These choices are most likely caused by confusion and are hard to interpret. We therefore decided to exclude these 235 participants from the analysis for this task following the standard procedure for this task (see Charness, Gneezy, & Imas, 2013, p. 50, for a detailed justification of the exclusion procedure) but they remained in our sample for all other analyses.

Potential for demand effects

We define the potential for demand effects as the proportion of participants who correctly guess the research hypothesis of the study (i.e., identify the correct independent and dependent variables and correctly state the direction of their relation). We used two measures to elicit whether participants correctly guessed the research hypothesis: an unstructured measure (open-form question) and a structured measure (with pre-defined answers). Both these measures were elicited after participants had completed the decision part of the study.

More precisely, we first asked the participants a very open, free-form question: "*What is your guess about the aim of this study? We are interested in your personal opinion*". The variable was scored 1 if participants mentioned anything related to power or risk-taking and 0 otherwise.

Table 2
Correlation matrix.

Variables	Mean	Std.	1	2	3	4	5	6	7	8	9
1. Age	35.80	10.75	_								
2. Gender	0.57	0.50	−0.14**	_							
3. PP	0.48	0.50	−0.01	−0.01	_						
4. Power	0.50	0.50	−0.01	−0.00	0.00	_					
5. Manipulation check	0.50	0.50	0.05*	−0.02	−0.01	0.01	_				
6. Demand_SM	0.29	0.45	0.02	−0.08**	0.22**	0.04	0.21**	_			
7. Demand_UM	0.07	0.25	−0.04	0.03	0.17**	0.00	0.07**	0.24**	_		
8. Self-reported risk-taking	4.36	1.63	−0.16**	0.24**	0.01	0.05*	−0.02	−0.15**	−0.05	_	
9. Bart risk-taking	18.76	11.75	0.03	0.06*	0.02	0.01	0.01	0.09**	0.07**	−0.01	_
10. Falk risk-taking	2.64	1.43	−0.13**	0.09**	0.00	0.00	−0.04	0.00	−0.02	0.23**	0.10**

Note. Gender dummy coded as 0 = male and 1 = female; PP coded as 0 = power priming and 1 = resource allocation; Power coded as 0 = low-power and 1 = high-power; Manipulation check coded as 0 = without manipulation check and 1 with manipulation check; Demand_SM refers to the structured measure of potential for demand effects; Demand_UM refers to unstructured measure of potential for demand effects.
* $p < 0.01$.
** $p < 0.05$.

We did not incentivize the answers. This unstructured measure (UM) of the potential for demand effects is very conservative in the sense that it puts no pressure on participants to think hard about their answer and people were free to respond whatever they wanted (including evasive or neutral answers such as "I don't know"). We therefore see the data collected with this first measure as a lower bound for the true potential for demand effects, because it is likely that not all participants who (at least intuitively) understood the hypothesis actually also reported it.

Second, participants were faced with a multiple-choice question. They were asked to identify the relation investigated in the study. They had to select one single relation among 18 possible options presented to them. Examples of suggested relations were: "The relation between stereotype and self-control", "The relation between power and corruption", or "The relation between power and risk-taking". Participants were informed that if they selected the correct relation, they would gain a bonus of $0.1. Then, participants were asked: "Please tell us how you think the relation is? What could be the hypothesis (e.g., I think X increases/decreases Y)." Their answers were coded as 1 if participants mentioned the correct hypothesis, "*power increases risk-taking*" and 0 otherwise). This structured measure (SM) of the potential for demand effects is more "invasive" in the sense that the monetary incentives might induce participants to think much harder about the study hypothesis than they would have done otherwise. Moreover, the structured question also forces the participants to think about the study in terms of a hypothesis. We therefore see the data collected with this second method as an upper bound for the true potential for demand effects, because the method might not only have induced correct answers from those who had actively thought about the hypothesis of the study before, but also from those whose understanding was rather unconscious.

Results

Potential for demand effects

Correlations and descriptive statistics are presented in Table 2. We first focus on the impact of different power manipulations and the use of manipulation checks on our two measures for the potential for demand effects. As expected, overall more participants correctly guessed the study hypothesis when we elicited the potential for demand effects with the structured measure (SM: $M = 0.29$, $SD = 0.45$) than when we used the unstructured measure (UM: $M = 0.07$, $SD = 0.25$). We show how the different power manipulations (RA vs. PP) and the presence of a manipulation check influence the potential for demand effects in Table 3. The table displays OLS estimations in which we regress our two measures of the potential for demand effects (Panel A: SM, Panel B: UM) on indicator variables for the PP manipulation and the presence of a manipulation check. Both dependent variables are binary measures so that the estimated coefficients reported in the table directly correspond to (changes in) the percentage of correct guesses.

Column (1) reports the main effect of the power manipulation method in that it compares the potential for demand effects across the two conditions with RA and the two conditions with PP. In line with Hypothesis 1, we observe that using the PP method significantly increased the potential for demand effects (SM: $M = 0.39$, $SD = 0.49$, $p < 0.001$/UM: $M = 0.11$, $SD = 0.32$) in comparison to the PP method (SM: $M = 0.20$, $SD = 0.40$/UM: $M = 0.03$, $SD = 0.17$, $p < 0.001$).[6]

Column (2) estimates the main effect of the manipulation check in that it compares across the two conditions without and the two conditions with a manipulation check. As predicted in Hypothesis 2, we find that the presence of a manipulation check significantly increased the potential for demand effects (SM: $M = 0.39$, $SD = 0.49$/UM: $M = 0.09$, $SD = 0.28$) in comparison to the absence of a manipulation check (SM: $M = 0.20$, $SD = 0.40$), $p < 0.001$/UM: $M = 0.05$, $SD = 0.22$), $p = 0.006$).[7]

Column (3) finally provides the full model in which we include both indicator variables and their interaction as predictors. This estimation allows us to compare the potential for demand effects across all four of our conditions. Fig. 3 graphically represents the corresponding condition-specific means (percentage of correct guesses) for our two measures of the potential for demand effects (Panel A: SM, Panel B: UM). As expected, participants in the least demand-prone condition (RA without manipulation check) correctly guessed the hypothesis of the study less often (SM:

[6] One might argue that power primes are poorly suited for MTurk users, because those participants have been over-exposed to the method (Hauser, Paolacci, & Chandler, 2019; Zhou & Fishbach, 2016). In a study on power primes in mTurk, Rinderknecht (2019) notes that "*41% of participants reported encountering such primes on at least a weekly basis, and 81% of participants reported encounters on at least a monthly basis.*" Therefore, in our study, we asked participants in the control questions whether they had been confronted with such episodic recall tasks before. Among all participants in the power priming condition, 25% self-reported that they had encountered such tasks before. We thus re-analyzed our data by splitting the sample in the priming conditions into a "naïve" sample (75%) and an experienced sample (25%). When redoing the analyses based on only the 75% of the participants who were naïve, our results remained robust. We observed that using the RA method reduced the potential for demand effects (SM: $M = 0.19$, $SD = 0.40$/UM: $M = 0.03$, $SD = 0.17$) in comparison to the PP method (SM: $M = 0.37$, $SD = 0.48$, $F(1,1434) = 60.13$, $p < 0.001$; UM: $M = 0.09$, $SD = 0.28$, $F(1,1434) = 25.90$, $p < 0.001$). We now state that the results remain unchanged when excluding the 25% non-naïve participants (with respect to the PP). Moreover, comparing the percentages of correct guesses among the naïve and non-naïve participants, the results show that naïve participants guessed the correct hypothesis 37% of the time and the non-naïve participants 47% of the time ($F(1,780) = 5.59$, $p = 0.02$). The 37% of correct guesses clearly indicates that even when not being familiar with the recall task, people are able to guess the aim of the study correctly. For RA method the correct guesses are about 20% of the time and the difference between the correct guesses in the PP and RA methods for naïve participants is also significant ($F(1,1434) = 60.13$, $p < 0.001$). As the results remain unchanged when excluding the 25% non-naïve participants (with respect to the PP) we can conclude that, our results are independent of the participant's prior experience with episodic recall tasks.

[7] Adding control variables (gender, age, and ethnicity) did not change the results. Please see *Table 3*.

Table 3
OLS regression analyses estimating the potential for demand effects.

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Structured Measure (SM).						
PP	0.20*** (0.02)		0.26** (0.03)	0.19** (0.02)		0.25** (0.03)
Manipulation check		0.19** (0.02)	0.25** (0.03)		0.19** (0.02)	0.25** (0.03)
PP × Manipulation check			−0.12** (0.04)			−0.12** (0.04)
Control variables				Incl.	Incl.	Incl.
Constant	0.20** (0.01)	0.20** (0.01)	0.07** (0.01)	0.10 (0.14)	0.09 (0.18)	−0.04 (0.16)
R^2	0.047	0.044	0.096	0.062	0.063	0.11
Panel B: Unstructured Measure (UM).						
PP	0.08** (0.01)		0.10** (0.02)	0.08** (0.01)		0.10** (0.02)
Manipulation check		0.03** (0.01)	0.05** (0.01)		0.04** (0.01)	0.05** (0.01)
PP × Manipulation check			−0.02 (0.03)			−0.03 (0.03)
Control variables				Incl.	Incl.	Incl.
Constant	0.03** (0.01)	0.05** (0.01)	0.01 (0.00)	0.15 (0.15)	0.17 (0.15)	0.12 (0.15)
R^2	0.028	0.005	0.033	0.036	0.016	0.042

Note. Robust standard errors are presented in parentheses. PP coded as 0 = resource allocation and 1 = power priming; Manipulation check coded as 0 = without manipulation check and 1 with manipulation check. Control variables included age, gender dummy coded as 1 = female and 0 = male, and six individual dummy variables for the ethnicities of American Indian (or Alaska Native), Asian, Black (or African American), White, Native Hawaiian, and Other.
** $p < 0.01$.
* $p < 0.05$.

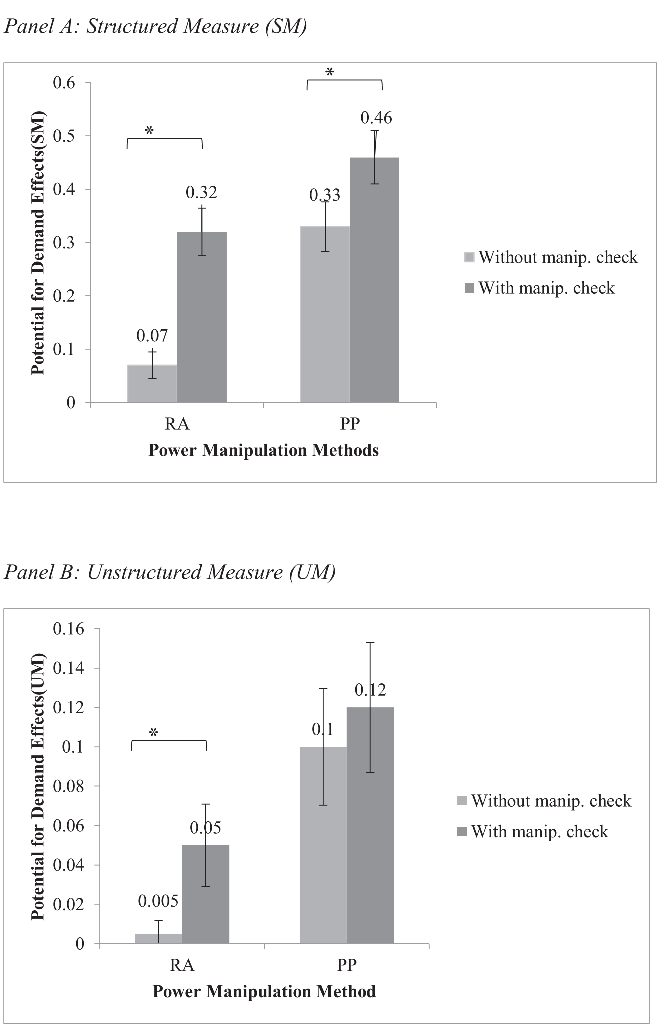

Fig. 3. The effect of power manipulation method and manipulation check on the potential for demand effects using a structured measure (SM) and an unstructured measure (UM). PP = power priming, RA = resource allocation; With manip. check = with manipulation check, Without manip. check = without manipulation check. (*) indicates statistically significant difference ($p < 0.01$).

$M = 0.07$, $SD = 0.26$/UM: $M = 0.005$, $SD = 0.07$) than participants in any of the other conditions. In line with Hypothesis 1, the addition of a manipulation check (RA with manipulation check) significantly increased the potential for demand effects (SM: $M = 0.32$, $SD = 0.47$, $p < 0.001$/UM: $M = 0.05$, $SD = 0.22$, $p = 0.006$). Similarly, and consistent with Hypothesis 2, switching to power priming (PP without manipulation check) also leads to a significantly higher potential for demand effects (SM: $M = 0.33$, $SD = 0.47$, $p < 0.001$/UM: $M = 0.10$, $SD = 0.30$, $p < 0.001$).

Although our theorizing has not produced hypotheses regarding the relative importance and the interaction of the two sources for the demand effect, Column (3) and Fig. 3 also allow us to assess these aspects in our setting. It turns out that both our measures indicate that the potential for demand effects is highest in the condition that combines the PP power manipulation with a manipulation check (SM: $M = 0.46$, $SD = 0.50$/UM: $M = 0.12$, $SD = 0.33$). However, the relative contribution of the two sources somewhat differs across the measures.

The analysis of the structured measure indicates that the individual impacts of the PP condition and the manipulation check are approximately equal in size ($p = 0.79$), whereas their joint presence (PP with manipulation check) produces demand effects potential which is significantly larger than in any other condition ($p < 0.001$, for all three comparisons).

When analyzing the unstructured measure, we find that switching from the RA manipulation to the PP manipulation has a larger impact on the potential for demand effects than adding a manipulation check to the RA manipulation ($p = 0.004$). Combining the two sources does then not lead to a further increase in the demand effects potential relative to using the PP manipulation alone ($p = 0.20$), but it does shift up the demand effects potential relative to using RA in combination with a manipulation check ($p < 0.001$).

Risk-taking

In this second part of our Results section, we turn our attention to the impact of our different conditions on the link between power and risk-taking. As explained in the Method section, we used three measures for risk-taking all of which are commonly used in different literatures and which we thought would be differentially susceptible to demand. The three measures are obviously related (see a full representation of the correlations in Table 2), but they do not necessarily capture the exact same aspect of risk-taking, so that treating all measures as the same dependent variable does not seem appropriate. For this reason, we analyze each measure separately.

As a first step, we investigate whether we are able to replicate the positive impact of power on risk-taking documented in the previous

Table 4
OLS regression analyses estimating risk-taking.

Variables	(1) Self-report risk-taking	(2) Bart risk-taking	(3) Lottery risk-taking	(4) Self-report risk-taking	(5) Bart risk-taking	(6) Lottery risk-taking
Power	0.17* (0.08)	0.27 (0.59)	0.01 (0.08)			
RA without Manipulation check				4.32** (0.11)	18.96** (0.84)	2.77** (0.10)
RA with Manipulation check				4.25** (0.12)	17.81** (0.73)	2.61** (0.10)
PP without Manipulation check				4.32** (0.12)	18.41** (0.77)	2.57** (0.11)
PP with Manipulation check				4.22** (0.11)	19.38** (0.80)	2.59** (0.11)
RA without Manipulation check × Power				0.05(0.16)	−0.34 (1.25)	−0.10 (0.16)
RA with Manipulation check × Power				0.20 (0.16)	1.12 (1.07)	−0.10 (0.15)
PP without Manipulation check × Power				0.25 (0.16)	0.08 (1.12)	0.22 (0.15)
PP with Manipulation check × Power				0.18 (0.16)	0.18 (1.25)	0.05 (0.16)
Constant	4.28** (0.06)	18.63** (0.39)	2.64** (0.05)	–	–	–
R^2	0.003	0.0001	0.000	0.878	0.719	0.773

Note. Robust standard errors are presented in parentheses. Power dummy coded as 0 = low-power and 1 = high-power; PP = power priming; RA = resource allocation.
** $p < 0.01$.
* $p < 0.05$.

literature. Columns (1) to (3) of Table 4 report OLS estimations in which we regress each of our three risk-taking measures on an indicator variable for power.[8] These regressions pool the data of all our power manipulations. We only find very weak support for an effect of power on risk-taking. When using the non-consequential measure of risk-taking (self-report variable ranging from 1 (unwilling to take risks) to 7 (fully prepared to take risks), see Column (1)), we find a statistically significant, positive effect of power on risk-taking ($p = 0.04$), but the effect size is very small (an increase of merely 4%). We do not find evidence for an effect of power on risk-taking for our consequential measures of risk-taking (BART, Column (2), $p = 0.65$, and lottery risk-taking task, Column (3), $p = 0.87$). There is therefore only very limited support for our Hypothesis 3a and the support only comes from the non-consequential measure that we deem as being rather unsuitable to measure behavior in a reliable manner.

Hypothesis 3b predicted that demand-prone methods would produce larger effects of power on risk-taking. Columns (4) to (6) of Table 4 provide a first analysis of this hypothesis. We report OLS estimations (without constants) in which we regress our three measures of risk-taking on indicator variables for each of our four conditions and interactions of these treatment variables with the indicator variable for power. The interaction terms provide direct measures of the condition-specific impact of power on risk-taking. Fig. 4 provides a graphical representation of these results.

We first focus on the non-consequential measure of risk-taking (see Column (4) in Table 4 and Panel A in Fig. 4). Hypothesis 3b suggests that we should see a larger impact of power on risk-taking in those conditions that use power priming and/or a manipulation check. Directionally, we find some support for this hypothesis. In our power priming conditions we find effect sizes of 6% (PP with manipulation check, $p = 0.13$) and 4% (PP without manipulation check, $p = 0.25$), and in the RA condition with manipulation check the effect size is 5% ($p = 0.22$). In the RA condition without a manipulation check, in contrast, the effect size is only 1% ($p = 0.76$). However, as the p-values reported above demonstrate, none of the condition-specific effects of power is significantly different from zero (and the effects do not significantly differ across conditions).[9]

Nevertheless, it seems clear that the (small) main effect of power on non-consequentially measured risk-taking reported in Column (1) is mostly driven by the more demand-prone conditions. In fact, if we exclude the data from the condition with the lowest potential for demand effects (RA without manipulation check) from the regression in Column (1), both the effect size and the significance become stronger (5%, $p = 0.03$). Removing the data from any other condition, in contrast, renders the effect of power non-significant.[10]

The estimations reported in Columns (5) and (6) of Table 4 (see also Panels B and C in Fig. 4) show that the absence of an effect of power on risk-taking for our consequential measures of risk-taking is fully confirmed when each single condition is considered.

Taken together, this first analysis of Hypothesis 3b yields no strong support—not even in the case of our non-consequential measure of risk-taking where we expected the effect to be most pronounced (see the discussion of Hypothesis 3b above).

As a second alternative attempt and an arguably more directive test of Hypothesis 3b, we use an instrumental variables (IV) approach. The aim of this approach is to regress risk-taking on an indicator variable for power, a continuous variable measuring the potential for demand effects, and the interaction of these variables. A significant interaction effect in such an estimation would provide direct support for our hypothesis (in particular, when employing non-consequentially measured risk-taking as the dependent variable). Performing such an analysis by using the measured potential for demand effects in an OLS regression would be problematic because of endogeneity concerns. We therefore follow the approach suggested by Antonakis, Bendahan, Jacquart, and Lalive (2010) and recently laid out in detail by Sajons (2020) and use our exogenous treatment variations as instruments for the potential for demand effects and the interaction effect. Specifically, we perform a two-stage least squares estimation in which the first stage regresses the potential for demand effects and the interaction effect on indicator variables for using the PP manipulation, using a manipulation check, and their interaction. The second stage then uses the predicted values from the first stage as exogenous regressors to estimate the effects of interest. A first-stage F-statistic showed the strength of our instruments that passed the required critical value of 9.08 (Stock & Yogo, 2005).[11] Despite the non-significant relationship between power and risk-taking, the 2SLS regression diagnostics suggested that it was reasonable

[8] The specification reported in Table 4 is mathematically equivalent to a specification in which the risk-taking measure is regressed on indicator variables for power priming, the presence of manipulation check, power, the interaction of power priming and power, the interaction of the presence of a manipulation check and power, and the three-way interaction of power priming, the presence of a manipulation check and power. The advantage of our specification is that the impact of power on risk-taking in each of our four manipulations can be seen directly in the table. The alternative specification mentioned above yields identical results, but requires testing linear combinations of estimated coefficients to determine the condition-specific effects of power.

[9] The p-values for the pairwise comparisons are as follows: RA without manipulation check vs. RA with manipulation check: $p = 0.50$ / RA without manipulation check vs. PP without manipulation check: $p = 0.37$ / RA without manipulation check vs. PP with manipulation check: $p = 0.54$ / RA with manipulation check vs. PP without manipulation check: $p = 0.83$ / RA with manipulation check vs. PP with manipulation check: $p = 0.94$ / PP without manipulation check vs. PP with manipulation check: $p = 0.77$.

[10] Removing one of the other three conditions leads to the following effect sizes and p-values: Removing RA with manipulation check: 4%, $p = 0.09$ / Removing PP without manipulation check: 3%, $p = 0.12$ / Removing PP with manipulation check: 4%, $p = 0.08$.

[11] The result of the F-test on the strength of the instruments for self-reported risk-taking is $F(5,1626) = 54.42$, $p < 0.001$ for the potential for the demand effects and is $F(5,1602) = 32.69$, $p < 0.001$ for its interaction with power; for Bart risk-taking is $F(5,1602) = 54.78$, $p < 0.001$ for the potential for the demand and is $F(5,1602) = 33.07$, $p < 0.001$ for its interaction with power; and for Lottery risk-taking is $F(5,1389) = 52.38$, $p < 0.001$ for the potential for the demand and is $F(5,1389) = 31.04$, $p < 0.001$ for its interaction with power.

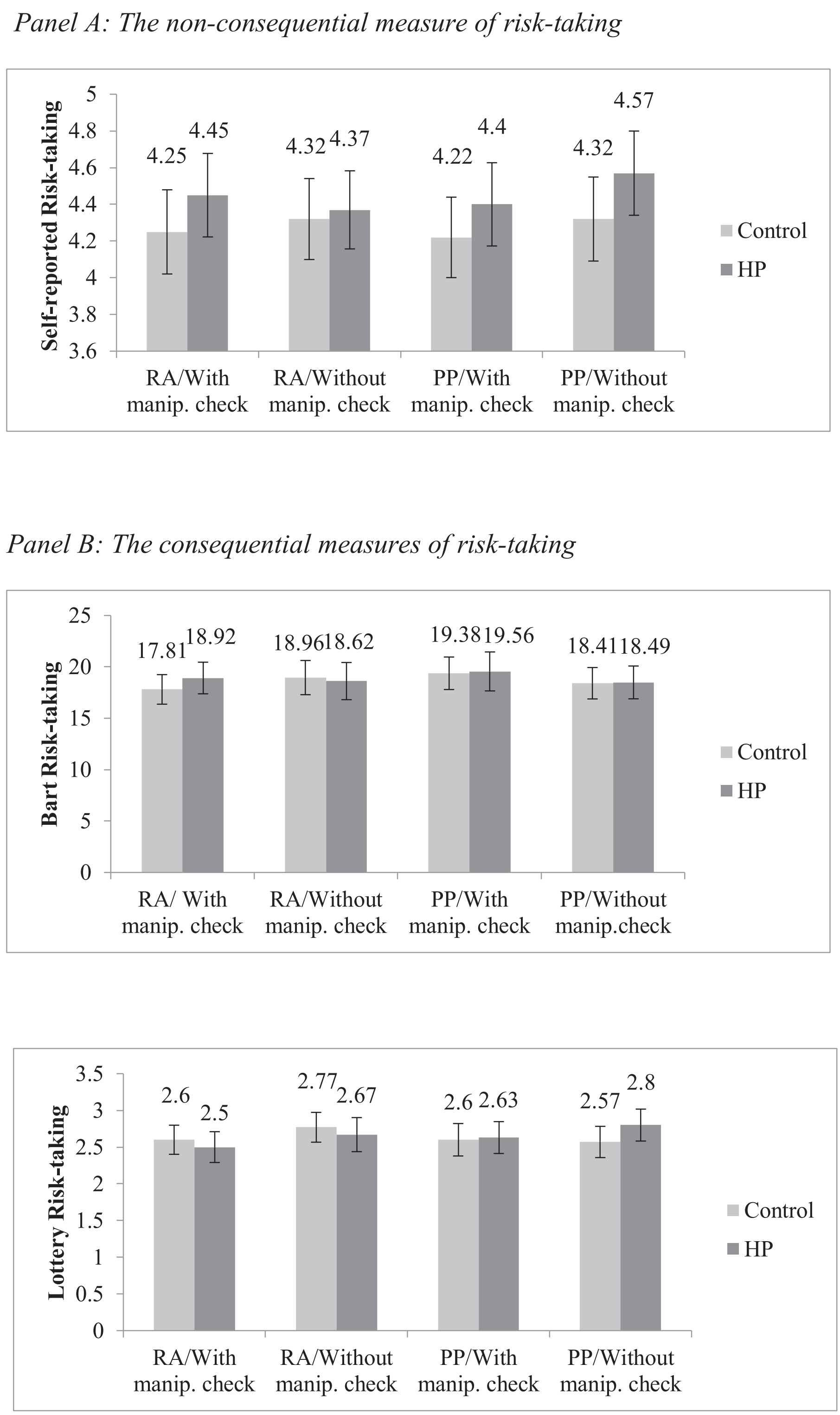

Fig. 4. The effect of power manipulation method, manipulation check, and power level on risk-taking. PP = power priming, RA = resource allocation; With manip. Check = with manipulation check, Without manip. Check = without manipulation check; HP = high-power.

to use these IVs (for self-reported measure: underidentification test, $\chi2(4) = 85.60$, $p < 0.001$, Sargan-Hansen test, $\chi2(3) = 1.51$, $p = 0.68$; for Bart risk-taking: underidentification test, $\chi2(4) = 85.45$, $p < 0.001$, Sargan-Hansen test, $\chi2(3) = 2.38$, $p = 0.50$; and for Falk risk-taking: underidentification test, $\chi2(4) = 82.30$, $p < 0.001$, Sargan-Hansen test, $\chi2(3) = 3.67$, $p = 0.30$). We report the results of those estimations for each of our three measures of risk-taking in Table 5. Given the previous analysis and the fact that we barely succeed in replicating the impact of power on risk-taking, it is not surprising that these results do not show support for Hypothesis 3.

Discussion

In this paper, we made a first step at addressing the important concern that using power priming (PP) might lead to demand-driven results. In an incentivized experimental design, we showed that using the PP manipulation method and/or adding a manipulation check significantly increased the potential for demand effects compared to using the resource allocation (RA) method. We used two measures to assess the potential for demand effects. Our unstructured measure (UM) indicates that the share of participants who correctly guess the hypothesis of the study increases from 0.5% to 12% when using PP in combination with a manipulation check compared to using RA with a manipulation check. One might be tempted to argue that these numbers imply that the potential for demand effects is rather low in general. However, it is important to keep in mind that these levels need to be regarded as a lower bound, because they most likely underestimate the true relevance of the problem. The UM was not incentivized and provided very little guidance on how to formulate what the aim of the study might have been. It is therefore likely that not all participants who were aware of the study hypothesis were identified by the UM. Our second measure, the structured measure (SM), indicates an increase from 7% to 46% in the potential for demand effects when moving from the RA without a manipulation check to the PP with a manipulation check. These numbers reveal that the magnitude of the problem might be much larger than indicated by the conservative UM. We acknowledge that the levels suggested by the SM should probably be seen as an upper bound, because the financial incentives and the prestructured answers might have induced some participants to think more actively and possibly in a different manner about the hypothesis of the study than in the absence of such a measure. However, it is important to emphasize that the SM does not inflate the measured potential for demand effects by construction, because it remains at a low level for the non demand-prone manipulation (RA without manipulation check). Taken together our results show that the concern that PP creates a large potential for demand effects cannot be dismissed.

We also investigated the extent to which the higher potential for demand effects created by demand-prone manipulations translates into the actual demand effects. For this purpose, we use a behavioral trait that the literature has frequently discussed: risk-taking. We elicited risk-taking using three different measures, two of which were incentivized, and one which was non-consequential. We find only limited evidence in support of our hypothesis that the demand-prone manipulations lead to upward biased effect sizes. The main reason is that we are barely able to replicate the positive effect of power on risk-taking in general. When we combine all of our four power manipulations, we find a significant main effect of power on the non-consequential measures of risk-taking. However, the overall effect size is very small, and we are therefore unable to find significant differences across manipulation types. We do not find any evidence for an effect of power on risk-taking when using the incentivized measures. Further analyses reveal that the significant main effect of power on the non-consequential measure of risk-taking is mostly driven by the demand-prone conditions, but this is admittedly not very strong evidence.

We understand that some people will argue that our data show that demand-prone manipulations are not a big issue after all. Even though power priming and manipulation checks create a big potential for demand effects (because many people correctly guess the hypothesis of the study), these methods do not ultimately lead to a substantial bias in the relation of interest. We strongly disagree with this view. Knowledge about the research hypothesis (potential for demand effects) would not always result in an aligned behavioral response to this knowledge (demand effect) (Nichols & Maner, 2008). For instance, it might be the case that people adjust their behavior in a specific way (e.g., knowing the hypothesis that men would tend to behave in a sexist way in a situation, some may try to actively counteract such a hypothesis and show either a neural or exactly an opposite behavior from what is expected). Such traces of demand characteristics are empirically much more difficult to test. Based on currently existing knowledge it is nearly impossible to predict under what circumstances a high potential for demand effects leads to actual demand effects and when not. Thus, when using methods that create a large potential for demand effects, researchers can never be sure whether their results are a true effect or demand effect. There are two solutions to this problem: either we explicitly test for each variable of interest whether demand effects are a problem or not, or we rely on methods that limit the potential for demand effects. We believe that the former solution makes no sense, because this approach would require many studies of the type that we report in this paper. For this reason, we advocate the use of methods that minimize the potential for demand effects. Based on these reflections, we would argue that the most important results in our paper are those that relate to the potential for demand effects, because they point to the source of the (potential) problem.

Despite the low implementation cost of the PP method, the potentially high threat of demand effects suggests that power researchers should use alternative methods to manipulate power, especially when they want to predict behavior of people with actual power. The RA method, suggested here, strongly reduces the potential for demand effects and has the advantage to induce actual power (i.e., having actual control over valued resources). This also increases construct validity of the power manipulation.

Note, however, that the RA method only produced a very low potential for demand effects when it was not followed by a manipulation check. This underlines the importance of paying attention to different elements of the experimental design that can reveal the hypothesis of the study to the participants (e.g., demand-prone method as well as the manipulation checks). In the RA method, the control over valued resources and the decision on the outcomes of others grant actual power to powerholders (as opposed to the subjective power in the priming method) and thus remove the need for a manipulation check. Because power is manipulated subjectively in the PP method, it is recommended to implement manipulation checks to examine the effect of the manipulation (Sigall & Mills, 1998). However, in such cases the researcher should either conduct the manipulation in a separate (pre-)study or add enough filler items to the manipulation check scale in order to conceal the research goal or topic (Ejelöv & Luke, 2020). Ideally, one should use an independent but comparable sample to check the effect of the manipulation (Fayant et al., 2017; Hauser et al., 2018; Kidd, 1976).

In our paper, we manipulated power by giving people actual control over valued resources. We assumed that the lottery tickets will be of value because they have an expected value that is ten times higher than the standard recruitment fee on Mturk.[12] We consciously did not allow allocators to take any ticket for themselves and thus eliminated the possibility of opportunistic behavior in the RA treatment to avoid a potential confound through income effects as well as adding more experimental control. If the allocators could keep lottery tickets for themselves, this would lead to differences in (expected) monetary payoffs before they enter the risk-taking task. These endogenous differences in endowments may affect behavior in the risk-taking tasks. Our manipulation avoids this issue. But there are research questions for which the opportunistic dimension of power might be central (e.g., to study the link between power and corruption). Designs targeted at such questions need to come up with solutions to deal with the above-mentioned threat of income effects.

The RA manipulation focused on the distribution of valued resources, which is a relevant source of power in many real-life situations. Yet, as discussed in the introduction other relevant sources of power exist as

[12] Amazon Mechanical Turk stipulates that participants can be recruited for $0.01 per assignment (retrieved from https://www.mturk.com/pricing). Therefore, we assume that having a ticket that offers a 1 in 10 chance to win $1 (an expected value that is ten times higher than the recruitment fee) and requires no further effort or work from the participant would be held as valuable.

Table 5
Two-stage least squares regression analyses estimating risk-taking.

Variables	(1) Self-reported risk-taking	(2) Bart risk-taking	(3) Lottery risk-taking
First stage regression of Demand effects (SM)			
PP	0.24** (0.03)	0.24** (0.03)	0.27** (0.04)
Manipulation check	0.22** (0.04)	0.22** (0.03)	0.26** (0.04)
Power	−0.01 (0.03)	−0.01 (0.03)	−0.004 (0.03)
Manipulation check × Power	0.05 (0.04)	0.06 (0.04)	0.08 (0.05)
PP × Power	0.04 (0.04)	0.04 (0.04)	0.03 (0.05)
Manipulation check × PP	−0.12** (0.04)	−0.12** (0.04)	−0.18** (0.05)
Constant	0.08** (0.02)	0.08** (0.02)	0.08** (0.02)
First stage regression of the interaction between Power and Demand effects (SM)			
PP	0.03 (0.02)	0.03 (0.02)	0.04* (0.02)
Manipulation check	0.03 (0.01)	0.03 (0.01)	0.04* (0.02)
Power	0.09** (0.02)	0.09** (0.02)	0.12** (0.02)
Manipulation check × Power	0.22** (0.03)	0.22** (0.03)	0.25** (0.03)
PP × Power	0.22** (0.03)	0.22** (0.03)	0.21** (0.03)
Manipulation check × PP	−0.05 (0.03)	−0.05 (0.03)	−0.08* (0.03)
Constant	−0.01 (0.01)	−0.01 (0.01)	−0.02* (0.01)
Second stage regression of Risk-taking			
Power level × Demand_SM	0.34 (0.58)	1.37 (4.47)	0.41 (0.55)
Power	0.07 (0.19)	−0.17 (1.45)	−0.10 (0.19)
Demand_SM	−0.21 (0.44)	0.31 (3.29)	−0.53 (0.42)
Constant	4.34** (0.14)	18.54** (1.02)	2.80** (0.13)
R^2	0.005	0.004	0.001

Note. Robust standard errors are presented in parentheses. PP coded as 0 = resource allocation and 1 = power priming; Power dummy coded as 1 = high-power and 0 = low-power; Manipulation check coded as 0 = without manipulation check and 1 = with manipulation check; Demand_SM refers to the structured measure of potential for demand effects. R^2 of the second stage of the IV estimation was calculated by taking the square of the correlation coefficient between the predicted and the true value of risk-taking measures (Bentler & Raykov, 2000). Because taking the model sum of squares (MSS) divided by the total sum of squares (TSS) in IV-regression when the endogenous variable and error terms are correlated is not informative and it can even be zero or negative (Wooldridge, 2009).
** $p < 0.01$.
* $p < 0.05$.

well (e.g., knowledge, status, etc., see Ocasio, Pozner, & Milner, 2020). We believe it is important that researchers are aware that power's different sources may lead to different forms of power with (potentially) different implications. Thus, power manipulations should be adapted to the forms and sources of power that are relevant for a particular research question. For example, using economic simulations where participants actually work with a powerful other can indeed be a promising way to manipulate reward/coercive power and using titles and high status roles can induce power concepts that are closer to such referent power or status. We therefore again emphasize that our manipulation is only one example of many possible manipulations of actual power. Future research might want to create a better understanding of power outcomes based on various types and proper manipulations of power.

Our research also highlights the importance of using consequential designs. Usually what we expect from behavioral studies is that they deal with individuals' behavior. But, how many researchers in the behavioral disciplines really use behavioral designs? Baumeister, Vohs, and Funder (2007) believe very few: the percentage of behavioral research has dropped from 80% in 1976 to less than 20% in 2006. What we observe is a decline in behavioral research which goes hand in hand with an increase in hypothetical and self-report procedures (Patterson, 2008; Patterson, Giles, & Teske, 2011). Sassenberg and Ditrich (2019) on their review of recent empirical research in social psychology pinpointed that to acquire larger statistical power, scholars increased the number of observations yet with the cost of using more self-report measures. The fields of management and leadership are also subject to similar critiques: actual behavior and choices in organizational settings are often not captured in various studies (Alvesson, 2020; Gottfredson, Wright, & Heaphy, 2020). Moreover, the underlying mechanisms are at times obscure (Alvesson & Einola, 2019; Blom & Alvesson, 2015). Questionnaires should not be fully relied upon to deduct the behavioral and psychological responses of the leaders and followers; instead attention should be focused on objective measures of actual behavior and using consequential designs (Eden, 2020; Fischer, Hambrick, Sajons, & Van Quaquebeke, 2020; Podsakoff & Podsakoff, 2019). Consequential experimental designs also matter to leadership in particular because they lead to robust findings that can have important implications for leaders in high stake positions (Antonakis, 2017).

Note that we are not against using self-reports and hypothetical designs. Self-report measures matter in many contexts, for example, when the researchers are interested in the emotion or perception of the individuals. However, they should not replace measuring actual behavior in real situations. Our results added to this discussion by pointing out that self-report measures of behavior are especially vulnerable to the effects of demand. We showed that consequential designs are more immune to the threat that demand-prone conditions pose.

In our study, we had a broad array of different risk-taking measures, yet there was no effect of power on consequential risk-taking. There are a number of possible explanations for why the effect of power on risk-taking was not replicated using the RA manipulation despite the large number of observations in this study and the correlations among the risk-taking measures. Based on a post-hoc power analysis (Faul, Erdfelder, Lang, & Buchner, 2007), assuming an alpha level of 0.05, and a small effect size of 0.2, by having 400 observations per cell we had a statistical power of 0.88 to detect the effect of power on risk-taking with our setting. Yet, we still could not replicate the effect. Scrutinizing the studies promoting the power/risk-taking effect, we noticed that many used self-report measures or hypothetical designs to capture risk-taking (e.g., see Studies 1 to 5, Anderson & Galinsky, 2006). Moreover, the main effect of power on risk-taking was not replicated in the context of consequential designs in several studies conducted by other researchers (see Studies 3 and 5, Jordan et al., 2011; Hiemer & Abele, 2012; Maner et al., 2007; Ronay & Von Hippel, 2010). Therefore, a possible explanation could be that there is no effect of power on risk-taking after all. We are also aware that our measures could not cleanly disentangle the pure effect of using a non-consequential measure. Future replication studies on power and risk-taking may want to clarify this state of affairs and consider the effect size in the presence and absence of consequential designs. For instance, one could think about adding a condition to our design in which participants are solely being asked in a hypothetical scenario how they would choose in the lottery risk-taking task (a pure hypothetical design with no consequences). This set-up would allow to compare the different outcomes directly in one study.

Our research addresses concerns raised about the potential demand characteristics in power research and the pervasive use of the PP method (Lonati et al., 2018; Sturm & Antonakis, 2015; Schaerer et al., 2018), as well as about the problems of using manipulation checks before measuring the dependent variable (Ejelöv & Luke, 2020; Fayant et al., 2017; Hauser et al., 2018). We also contribute to the literature discussing experimenter demand effects (de Quidt et al., 2018; de Quidt et al., 2019; Nichols & Maner, 2008; Zizzo, 2010). Our results highlight the importance of implementing rigorous research designs in order to derive ecologically valid conclusions valuable to both leadership scholars and practitioners. We suggest being prudent about revealing possible clues either through the manipulation method or via manipulation checks because if not, with a little help from demand characteristics, researchers may oftentimes find support for what they hypothesized even if the effect is not really there.

Funding

This work was supported by HEC Research Fund for PhD and Postdocs. This work was supported by HEC Research Fund for PhD and Postdocs awarded by HEC Lausanne, Switzerland.

References

Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, *53*, 63–70.

Alvesson, M. (2020). Upbeat leadership: A recipe for–or against–"successful" leadership studies. *The Leadership Quarterly*, *31*(6), 1–12.

Alvesson, M., & Einola, K. (2019). Warning for excessive positivity: Authentic leadership and other traps in leadership studies. *The Leadership Quarterly*, *30(4)*, 383–395.

Anderson, C., & Berdahl, J. L. (2002). The experience of power: Examining the effects of power on approach and inhibition tendencies. *Journal of Personality and Social Psychology*, *83*(6), 1362–1377.

Anderson, C., & Galinsky, A. D. (2006). Power, optimism, and risk taking. *European Journal of Social Psychology*, *36*(4), 511–536.

Antonakis, J. (2017). On doing better science: From thrill of discovery to policy implications. *The Leadership Quarterly*, *28*(1), 5–21.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, *21*(6), 1086–1120.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*(7), 462–479.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*(4), 396–403.

Bendahan, S., Zehnder, C., Pralong, F. P., & Antonakis, J. (2015). Leader corruption depends on power and testosterone. *The Leadership Quarterly*, *26*(2), 101–122.

Bentler, P. M., & Raykov, T. (2000). On measures of explained variance in nonrecursive structural equation models. *Journal of Applied Psychology*, *85*(1), 125–131.

Blom, M., & Alvesson, M. (2015). All-inclusive and all good: The hegemonic ambiguity of leadership. *Scandinavian Journal of Management*, *31(4)*, 480–492.

Briñol, P., Petty, R. E., & Stavraki, M. (2012). Power increases the reliance on first-impression thoughts. *Revista de Psicología Social*, *27*(3), 293–303.

Carney, D. R., Cuddy, A. J., & Yap, A. J. (2010). Power posing: Brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychological Science*, *21*(10), 1363–1368.

Charness, G., Gneezy, U., & Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization*, *87*, 43–51.

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, *81*(1), 1–8.

Chester, D., & Lasko, E. (2019). *Construct validation of experimental manipulations in social psychology: current practices and recommendations for the future Retrieved from . https://doi.org/10.31234/osf.io/t7ev9.*

Chivers, T. (2019). What's next for psychology's embattled field of social priming. *Nature*, *576*, 200–202.

Colquitt, J. A. (2008). From the editors: Publishing laboratory research in AMJ: A question of when, not if. *Academy of Management Journal*, *51*(4), 616–620.

Cook, T. D., Campbell, D. T., & Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, *9*(3), 522–550.

Doldor, E. (2017). From politically naïve to politically mature: Conceptualizing leaders' political maturation journey. *British Journal of Management*, *28*(4), 666–686.

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), 1–7.

Dubois, D., Rucker, D. D., & Galinsky, A. D. (2010). The accentuation bias: Money literally looms larger (and sometimes smaller) to the powerless. *Social Psychological and Personality Science*, *1*(3), 199–205.

Durgin, F. H., Baird, J. A., Greenburg, M., Russell, R., Shaughnessy, K., & Waymouth, S. (2009). Who is being deceived? The experimental demands of wearing a backpack. *Psychonomic Bulletin & Review*, *16*(5), 964–969.

Eden, D. (2020). The science of leadership: A journey from survey research to field experimentation. *The Leadership Quarterly*. https://doi.org/10.1016/j.leaqua.2020.101472 (in press).

Ejelöv, E., & Luke, T. J. (2020). "Rarely safe to assume": Evaluating the use and interpretation of manipulation checks in experimental social psychology. *Journal of Experimental Social Psychology*, *87*, 1–13.

Emerson, R. M. (1962). Power-dependence relations. *American Sociological Review*, *27*(1), 31–41.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Fayant, M. P., Sigall, H., Lemonnier, A., Retsin, E., & Alexopoulos, T. (2017). On the limitations of manipulation checks: An obstacle toward cumulative science. *International Review of Social Psychology*, *30*(1), 125–130.

Fischer, T., Hambrick, D. C., Sajons, G. B., & Van Quaquebeke, N. (2020). Beyond the ritualized use of questionnaires: Toward a science of actual behaviors and psychological states [Special issue]. *The Leadership Quarterly*, *31*(4).

Flynn, F. J., Gruenfeld, D., Molm, L. D., & Polzer, J. T. (2011). Social psychological perspectives on power in organizations. *Administrative Science Quarterly*, *56*(4), 495–500.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, *6*(3), 347–369.

French, J. R. P., & Raven, B. (1959). The bases of social power. In D. Cartwright (Ed.), *Studies in social power* (pp. 150–167). Ann Arbor, MI: University of Michigan Press.

Galinsky, A. D., Gruenfeld, D. H., & Magee, J. C. (2003). From power to action. *Journal of Personality and Social Psychology*, *85*(3), 453–466.

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*(6), 1450–1466.

Galinsky, A. D., Rucker, D. D., & Magee, J. C. (2015). Power: Past findings, present considerations, and future directions. In M. Mikulincer, P. R. Shaver, J. A. Simpson, & J. F. Dovidio (Eds.), *APA handbook of personality and social psychology. Vol. 3.* (pp. 421–460). Washington, DC: American Psychological Association.

Gervais, S. J., Guinote, A., Allen, J., & Slabu, L. (2013). Power increases situated creativity. *Social Influence*, *8*(4), 294–311.

Gilder, T. S., & Heerey, E. A. (2018). The role of experimenter belief in social priming. *Psychological Science*, *29*(3), 403–417.

Giurge, L. M., van Dijke, M., Zheng, M. X., & De Cremer, D. (2020). Does power corrupt the mind? The influence of power on moral reasoning and self-interested behavior. *The Leadership Quarterly* Retrieved from https://www.sciencedirect.com/science/article/pii/S1048984316301138 (in press).

Goldstein, N. J., & Hays, N. A. (2011). Illusory power transference: The vicarious experience of power. *Administrative Science Quarterly*, *56*(4), 593–621.

Gordon, R. D. (2002). Conceptualizing leadership with respect to its historical–contextual antecedents to power. *The Leadership Quarterly*, *13*(2), 151–167.

Gottfredson, R. K., Wright, S. L., & Heaphy, E. D. (2020). A critique of the leader-member exchange construct: Back to square one. *The Leadership Quarterly*, *31*(6), 1–17.

Gruenfeld, D. H., Inesi, M. E., Magee, J. C., & Galinsky, A. D. (2008). Power and the objectification of social targets. *Journal of Personality and Social Psychology*, *95*(1), 111–127.

Guinote, A. (2007a). Power and goal pursuit. *Personality and Social Psychology Bulletin*, *33*(8), 1076–1087.

Guinote, A. (2007b). Behavior variability and the situated focus theory of power. *European Review of Social Psychology*, *18*(1), 256–295.

Harris, C. R., Coburn, N., Rohrer, D., & Pashler, H. (2013). Two failures to replicate high-performance-goal priming effects. *PLoS One*, *8*(8), 1–9.

Hauser, D., Paolacci, G., & Chandler, J. (2019). Common concerns with MTurk as a participant Pool. In F. R. Kardes, P. M. Herr, & N. Schwarz (Eds.), *Handbook of research methods in consumer psychology* (pp. 319–336). New York, NY: Routledge.

Hauser, D. J., Ellsworth, P. C., & Gonzalez, R. (2018). Are manipulation checks necessary? *Frontiers in Psychology*, *9*(998), 1–10.

Hiemer, J., & Abele, A. E. (2012). High power = motivation? Low power = situation? The impact of power, power stability and power motivation on risk-taking. *Personality and Individual Differences*, *53*(4), 486–490.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644–1655.

Hu, M., Rucker, D. D., & Galinsky, A. D. (2016). From the immoral to the incorruptible: How prescriptive expectations turn the powerful into paragons of virtue. *Personality and Social Psychology Bulletin*, *42*(6), 826–837.

Jonas, K. J., Cesario, J., Alger, M., Bailey, A. H., Bombari, D., Carney, D., ... Jackson, B. (2017). Power poses–where do we stand? *Comprehensive Results in Social Psychology*, *2*(1), 139–141.

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*(4), 530–558.

Kahneman, D. (2012). A proposal to deal with questions about priming effects. Retrieved from https://www.nature.com/news/polopoly_fs/7.6716.1349271308!/suppinfoFile/Kahneman%20Letter.pdf (September 26).

Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review*, *110*(2), 265–284.

Kidd, R. F. (1976). Manipulation checks: Advantage or disadvantage? *Representative Research in Social Psychology*, *7*(2), 160–165.

Kühnen, U. (2010). Manipulation checks as manipulation: Another look at the ease-of-retrieval heuristic. *Personality and Social Psychology Bulletin*, *36*(1), 47–58.

Lammers, J., Galinsky, A. D., Gordijn, E. H., & Otten, S. (2012). Power increases social distance. *Social Psychological and Personality Science*, *3*(3), 282–290.

Lammers, J., Stapel, D. A., & Galinsky, A. D. (2010). Power increases hypocrisy: Moralizing in reasoning, immorality in behavior. *Psychological Science*, *21*(5), 737–744.

Lejuez, C. W., Aklin, W., Daughters, S., Zvolensky, M., Kahler, C., & Gwadz, M. (2007). Reliability and validity of the youth version of the balloon analogue risk task (BART–Y) in the assessment of risk-taking behavior among inner-city adolescents. *Journal of Clinical Child and Adolescent Psychology*, *36*(1), 106–111.

Lejuez, C. W., Aklin, W. M., Zvolensky, M. J., & Pedulla, C. M. (2003). Evaluation of the balloon analogue risk (BART) as a predictor of adolescent real-world risk-taking behaviors. *Journal of Adolescence*, *26*(4), 475–479.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., ... Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The balloon analogue risk task (BART). *Journal of Experimental Psychology: Applied*, *8*(2), 75–84.

Lewellyn, K. B., & Muller-Kahle, M. I. (2012). CEO power and risk taking: Evidence from the subprime lending industry. *Corporate Governance: An International Review*, *20*(3), 289–307.

Lonati, S., Quiroga, B. F., Zehnder, C., & Antonakis, J. (2018). On doing relevant and rigorous experiments: Review and recommendations. *Journal of Operations Management*, *64*, 19–40.

Magee, J. C., & Galinsky, A. D. (2008). 8 social hierarchy: The self-reinforcing nature of power and status. *Academy of Management Annals*, *2*(1), 351–398.

Magee, J. C., & Smith, P. K. (2013). The social distance theory of power. *Personality and Social Psychology Review*, *17*(2), 158–186.

Maner, J. K., Gailliot, M. T., Butz, D. A., & Peruche, B. M. (2007). Power, risk, and the status quo: Does power promote riskier or more conservative decision making? *Personality and Social Psychology Bulletin*, *33*(4), 451–462.

Nichols, A. L., & Maner, J. K. (2008). The good-subject effect: Investigating participant demand characteristics. *The Journal of General Psychology*, *135*(2), 151–166.

Ocasio, W., Pozner, J. E., & Milner, D. (2020). Varieties of political capital and power in organizations: A review and integrative framework. *Academy of Management Annals*, *14*(1), 303–338.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*(11), 776–783.

Orne, M. T. (1969). Demand characteristics and the concept of quasi-controls. In R. Rosenthal, & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 143–179). New York, NY: Academic Press.

Pannenberg, M. (2010). Risk attitudes and reservation wages of unemployed workers: Evidence from panel data. *Economics Letters*, *106*(3), 223–226.

Pashler, H., Rohrer, D., & Harris, C. R. (2013). Can the goal of honesty be primed? *Journal of Experimental Social Psychology*, *49*(6), 959–964.

Patterson, M. L. (2008). Back to social behavior: Mining the mundane. *Basic and Applied Social Psychology*, *30*(2), 93–101.

Patterson, M. L., Giles, H., & Teske, M. (2011). The decline of behavioral research? Examining language and communication journals. *Journal of Language and Social Psychology*, *30*(3), 326–340.

Perdue, B. C., & Summers, J. O. (1986). Checking the success of manipulations in marketing experiments. *Journal of Marketing Research*, *23*(4), 317–326.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, *63*, 539–569.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management*, *12*(4), 531–544.

Podsakoff, P. M., & Podsakoff, N. P. (2019). Experimental designs in management and leadership research: Strengths, limitations, and recommendations for improving publishability. *The Leadership Quarterly*, *30*(1), 11–33.

de Quidt, J., Haushofer, J., & Roth, C. (2018). Measuring and bounding experimenter demand. *American Economic Review*, *108*(11), 3266–3302.

de Quidt, J., Vesterlund, L., & Wilson, A. J. (2019). Experimenter demand effects. In A. Schram, & A. Ule (Eds.), *Handbook of research methods and applications in experimental economics* (pp. 384–400). Cheltenham, UK: Edward Elgar Publishing.

Ranehill, E., Dreber, A., Johannesson, M., Leiberg, S., Sul, S., & Weber, R. A. (2015). Assessing the robustness of power posing: No effect on hormones and risk tolerance in a large sample of men and women. *Psychological Science*, *26*(5), 653–656.

Raven, B. H. (1993). The bases of power: Origins and recent developments. *Journal of Social Issues*, *49*(4), 227–251.

Rinderknecht, R. G. (2019). Effects of participant displeasure on the social-psychological study of power on Amazon's mechanical Turk. *SAGE Open*, *9*(3), 1–13.

Ronay, R., & Von Hippel, W. (2010). Power, testosterone, and risk taking. *Journal of Behavioral Decision Making*, *23*(5), 473–482.

Rosnow, R. L., & Rosenthal, R. (1997). *People studying people: Artifacts and ethics in behavioral research*. New York, NY: Freeman.

Rucker, D. D., Dubois, D., & Galinsky, A. D. (2011). Generous paupers and stingy princes: Power drives consumer spending on self versus others. *Journal of Consumer Research*, *37*(6), 1015–1029.

Sajons, G. B. (2020). Estimating the causal effect of measured endogenous variables: A tutorial on experimentally randomized instrumental variables. *The Leadership Quarterly*, *31*(5), 1–17.

Sassenberg, K., & Ditrich, L. (2019). Research in social psychology changed between 2011 and 2016: Larger sample sizes, more self-report measures, and more online studies. *Advances in Methods and Practices in Psychological Science*, *2*(2), 107–114.

Schaerer, M., du Plessis, C., Yap, A. J., & Thau, S. (2018). Low power individuals in social power research: A quantitative review, theoretical framework, and empirical test. *Organizational Behavior and Human Decision Processes*, *149*, 73–96.

Schaerer, M., Lee, A. J., Galinsky, A. D., & Thau, S. (2018). Contextualizing social power research within organizational behavior. In D. L. Ferris, R. E. Johnson, & C. Sedikides (Eds.), *The self at work: Fundamental theory and research* (pp. 194–221). New York, NY: Routledge.

Schmid, P. C. (2018). Less power, greater conflict: Low power increases the experience of conflict in multiple goal settings. *Social Psychology*, *49*(1), 47–62.

Schmid, P. C., Kleiman, T., & Amodio, D. M. (2015). Power effects on cognitive control: Turning conflict into action. *Journal of Experimental Psychology: General*, *144*(3), 655–663.

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), 1–10.

Sigall, H., & Mills, J. (1998). Measures of independent variables and mediators are useful in social psychology experiments: But are they necessary? *Personality and Social Psychology Review*, *2*(3), 218–226.

Smith, P. K., & Hofmann, W. (2016). Power in everyday life. *Proceedings of the National Academy of Sciences*, *113*(36), 10043–10048.

Smith, P. K., Jostmann, N. B., Galinsky, A. D., & van Dijk, W. W. (2008). Lacking power impairs executive functions. *Psychological Science*, *19*(5), 441–447.

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*(4), 578–596.

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. W. Andrews, & J. H. Stock (Eds.), *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg* (pp. 80–108). Cambridge: Cambridge University Press.

Sturm, R. E., & Antonakis, J. (2015). Interpersonal power: A review, critique, and research agenda. *Journal of Management*, *41*(1), 136–163.

Tost, L. P. (2015). When, why, and how do powerholders "feel the power"? Examining the links between structural and psychological power and reviving the connection between power and responsibility. *Research in Organizational Behavior*, *35*, 29–56.

Tost, L. P., Gino, F., & Larrick, R. P. (2012). Power, competitiveness, and advice taking: Why the powerful don't listen. *Organizational Behavior and Human Decision Processes*, *117*(1), 53–65.

Tost, L. P., Gino, F., & Larrick, R. P. (2013). When power makes others speechless: The negative impact of leader power on team performance. *Academy of Management Journal*, *56*(5), 1465–1486.

Tost, L. P., & Johnson, H. H. (2019). The prosocial side of power: How structural power over subordinates can promote social responsibility. *Organizational Behavior and Human Decision Processes*, *152*, 25–46.

de Wit, F. R., Scheepers, D., Ellemers, N., Sassenberg, K., & Scholl, A. (2017). Whether power holders construe their power as responsibility or opportunity influences their tendency to take advice from others. *Journal of Organizational Behavior*, *38*(7), 923–949.

Wooldridge, J. M. (2009). *Introductory econometrics: A modern approach*. Mason, OH: South-Western Cengage Learning.

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology*, *111*(4), 493.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*(1), 75–98.