

Low Hopes, High Expectations: Expectancy Effects and the Replicability of Behavioral Experiments

Olivier Klein¹, Stéphane Doyen^{1,2}, Christophe Leys¹,
Pedro A. Magalhães de Saldanha da Gama¹, Sarah Miller¹,
Laurence Questienne¹, and Axel Cleeremans^{1,2}

¹Université Libre de Bruxelles, Belgium, and ²Fonds de la Recherche Scientifique FRS-FNRS, Belgium

Perspectives on Psychological Science
7(6) 572–584

© The Author(s) 2012

Reprints and permission:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/1745691612463704

http://pps.sagepub.com



Abstract

This article revisits two classical issues in experimental methodology: experimenter bias and demand characteristics. We report a content analysis of the method section of experiments reported in two psychology journals (*Psychological Science* and the *Journal of Personality and Social Psychology*), focusing on aspects of the procedure associated with these two phenomena, such as mention of the presence of the experimenter, suspicion probing, and handling of deception. We note that such information is very often absent, which prevents observers from gauging the extent to which such factors influence the results. We consider the reasons that may explain this omission, including the automatization of psychology experiments, the evolution of research topics, and, most important, a view of research participants as passive receptacles of stimuli. Using a situated social cognition perspective, we emphasize the importance of integrating the social context of experiments in the explanation of psychological phenomena. We illustrate this argument via a controversy on stereotype-based behavioral priming effects.

Keywords

expectancy, demand characteristics, experimenter bias, replicability, methodology

The application of what we think of as the scientific method has not simplified human behavior. It has perhaps shown us more clearly just how complex it is.

—Robert Rosenthal and Ralph L. Rosnow
(1969/2009, p. 3)

All psychology experiments are fundamentally social endeavors that necessarily involve cooperation between researchers and participants. They enter the experiment with different motivations, expectations, and emotions. Such dispositions, far from being irrelevant to the outcomes of a psychological study, have in fact long been known to influence its results. Indeed, on the one hand, the role that participants' knowledge of the study's purpose or hypotheses may play in influencing the outcome has preoccupied researchers since the very early days of scientific psychology.¹ On the other hand, as the (in)famous case of Clever Hans (Despret, 2004; Pfungst, 1911/1965) so strikingly illustrated, experimenters' expectations are also prone to unwittingly influence a study's results. Robert Rosenthal (1975/2009) addressed this potential source of bias in a now classic line of research carried out mainly in the 1960s.

Since these seminal contributions, however, interest in such issues has steadily declined. Our main goal in this article is to document this decline and to reflect both upon its causes and on its consequences, hoping that this endeavor will address increasing concern about what has been dubbed the “replicability crisis”—the focus of this issue (cf. also, Pashler, Coburn, & Harris, in press; Young, 2012). Though in general there may be many reasons why a study fails to replicate (e.g., lack of statistical power; Chase & Chase, 1976; J. Cohen, 1962), we suggest that the central issue is a widespread failure to acknowledge, recognize, and document the circumstances and social context in which an experiment takes place. We support this claim through a brief content analysis of the method sections of experimental studies reported in two journals: *Psychological Science* and the *Journal of Personality and Social Psychology (JPSP)*. We show that information regarding the factors likely to influence either experimenters' expectations or demand characteristics is usually very sparse or simply

Corresponding Author:

Olivier Klein Université Libre de Bruxelles, Social Psychology Unit, 50 av. Roosevelt CP 122 Brussels, Brussels Capital District 1050, Belgium
E-mail: oklein@ulb.ac.be

lacking. We consider some of the factors that may explain the declining interest for these issues and suggest that they demand to be appraised in the context in a more general approach building on situated cognition. We then consider the relevance of this approach for the replicability crisis and conclude with specific recommendations on how to deal with experimenter bias.

Psychological Experimentation As a Social Endeavor

In his most influential article, titled “The Social Psychology of the Psychology Experiment,” Orne (1962) noted that psychologists had, perhaps out of anxiety to establish psychology as a genuine scientific discipline, espoused the experimental model of physics to design and carry out their own experiments. In adopting this holier-than-thou attitude, however, psychologists also paradoxically negated the enormous differences between their object of study—human beings—and that of physics:

(...) the use of such a model with animal or human subjects leads to the problem that the subject of the experiment is assumed at least implicitly to be a passive responder to stimuli (...). Further, in this type of model the experimental stimuli themselves are usually rigorously defined in terms of what is done to the subject. (Orne, 1962, p. 776)

Orne (1962), by contrast, insisted that one should focus on what the subject does. He suggested that aspects of the “experimental setting” may influence the subject’s reaction to the well-defined stimuli created by the investigator. Specifically, he proposed that participants in psychology experiments may be motivated to play the role of “good subjects.” This suggests that participants continuously and consciously attempt to reconstruct the experimenter’s hypotheses based on available cues, which Orne dubbed *demand characteristics*: “The demand characteristics of the situation help define the role of ‘good experimental subject’ and the responses of the subject are a function of the role that is created” (p. 779).

Thus, Orne argued, rather than being wholly passive, the experimental subject actually has a “real stake” in the success of the experiment. Such demand characteristics are unavoidable as they are part and parcel of the experimental procedure. Their influence can be subjected to empirical tests and have been demonstrated (see Nichols & Maner, 2008; Weber & Cook, 1972).

Let us now turn to the twin (and sometimes incestuous) brother of demand characteristics, that is, experimenter bias. Although as we already noted, this issue has preoccupied experimental psychologists since the very early days of the discipline, it became the object of renewed attention with Rosenthal’s work. Simply put, experimenters demonstrably exert a variety of different, unwanted effects on the outcome

of their experiments, thus exhibiting a form of unconscious bias. Rosenthal considered several sources of such bias (see Rosenthal, 1969/2009 for a summary). However, his main concern was about the influence of experimenters’ expectations on participants’ behavior. Thus, an experimenter who holds specific hypotheses regarding the effect of the independent variable(s) on behavior may unwittingly influence participants’ performance. Rosenthal (1969/2009) reviewed numerous studies on this topic and documented the prevalence of such effects. These influences seem to affect a host of different measures, such as self-reports, psychophysical judgments, interpretation of inkblots, and reaction times.

Relationships Between the Two Phenomena

Naturally, experimenters’ expectations and demand characteristics are not the only sources of “bias” in experimental psychology. For instance, the experimenter’s gender, his or her attire and demeanor, or the time and the setting where the experiment takes place have all been found to influence the results of behavioral experiments, though each of these factors are usually irrelevant to theoretical predictions (Rosenthal, 1975/2009). Such effects are presumably uniform across experimental conditions and therefore only contribute to increase the error term and thus reduce the power of the experiment. By contrast, demand effects and experimenters’ expectations influence the differences between conditions and may thus constitute a source of Type I error. Note also that these two phenomena are closely associated. The experimenter’s behavior is one of the main sources of demand characteristics and may be used by the participant to infer the behavior of a “good subject.” Nonetheless, the impact of experimenter’s expectations on the participant’s behavior is not necessarily mediated by conscious inference of the hypotheses. It may, for example, result from nonverbal behavior (e.g., Harris & Rosenthal, 1985; Klein & Snyder, 2003), such as unconscious behavioral mimicry (for a review, see Chartrand, Maddux, & Lakin, 2005).

From a methodological perspective, the two types of biases (i.e., those affecting the differences between conditions and error variance respectively) may be related. Indeed, a common response to lack of power (Type II error) consists in increasing sample size. But this practice may result in detecting (small) effects of little interest to the researcher, including those that derive from the experimenters’ expectations or demand characteristics rather than from the process of interest to the researcher.

What Do We Know About the Social Context of Psychology Experiments? An Archival Study

Experimenter bias and demand characteristics are just as much of a substantial issue today as they were when Rosenthal and

Orne first documented them. One may therefore wonder, “How are these crucial issues addressed in today’s research?” and “Have there been notable changes in the interest for these issues since the 1960s?” In an effort to provide answers to these questions, we computed the number of occurrences of words related to these two topics in the *JPSP* between 1965 and 2010. We chose this journal because it has been a primary outlet for social psychological research (in which experimenter expectancy effects may be more likely to occur) since the publication of Rosenthal’s book. We counted the number of occurrences of words related to experimenter bias and demand characteristics (see the list below Fig. 1) as reported by the Google Scholar database. As can be seen, frequencies have peaked in the 1970s before returning to a relatively low level in the 1980s. The pattern for experimenter bias and demand characteristics is relatively similar ($r = .57$). However, the frequency of words related to experimenter bias has been generally much lower since the early 1980s. Thus, based on these superficial indexes at least, it seems that experimenter bias and the impact of demand characteristics are more rarely acknowledged or simply neglected by contemporary psychologists than by their predecessors.

However, a failure to explicitly mention these words in the body of an article does not unequivocally demonstrate that the corresponding issues have been ignored from the design of the

experiments. To explore this issue more deeply, we thus set out to conduct a survey of the experimental studies published in two journals: one covering the whole field of psychology—*Psychological Science* (Vol. 16, 2005)—and the other covering only social psychology—the *JPSP* (Vol. 101, 2011).

We chose these two journals because they are the most important empirical periodicals in their respective fields (according to Thompson-Reuter’s Journal Citation Reports, 2011).²

Each study was coded individually and constituted our unit of analysis. Studies were considered for further analysis if (a) they included at least one experimentally manipulated variable, (b) they involved human participants older than 3 (as we did not expect demand characteristics to influence younger subjects), and (c) one of the focal dependent variables involved behavioral measures. Altogether, 176 studies met these criteria in *JPSP* and 170 met them in *Psychological Science*. For each article, we assessed the presence of information relevant to experimenters’ expectations and the impact of demand characteristics. A fraction of the studies (33 for *JPSP* and 52 for *Psychological Science*) were coded by a second judge to check the reliability of the coding. Disagreements were resolved by a third judge.

We focused on several points. First, we assessed whether the article provided information regarding the presence of the experimenter during the session (to perform this analysis, we

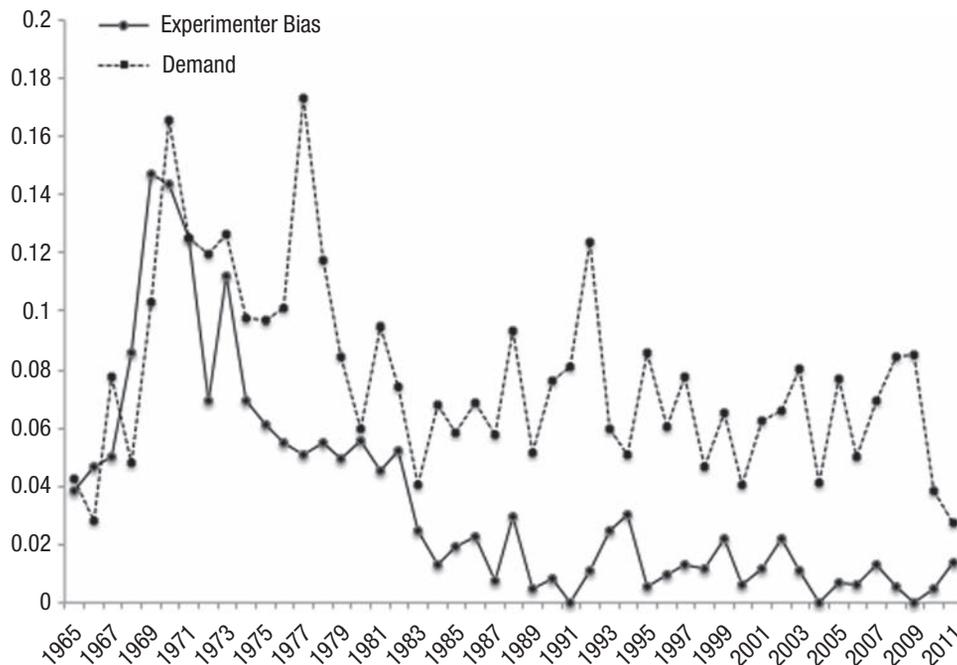


Fig. 1. Occurrence of words related to “demand characteristics” and “experimenter’s bias” in the *Journal of Personality and Social Psychology*.

Note: Each point represents the frequency of the relevant expressions divided by the total number of papers published during the relevant year in the *Journal of Personality and Social Psychology* as reported by Google Scholar (<http://scholar.google.com>). For the “experimenter bias,” the following expressions were coded: “experimenter bias,” “experimenter effects,” “experimenter expectancy,” “experimenter’s expectancy,” “experimenter expectation,” “experimenter’s expectation,” as well as the plural forms of these expressions. For “demand,” we coded “experimental demand” and “demand characteristics.”

first excluded online studies). As can be seen in Figure 2, the majority of the studies we considered fail to offer sufficient information to determine whether this was the case or not.

Second, we examined whether information was available regarding how the study had been presented to participants. Thus, for each study, we coded the presence or absence of such information, coding “yes” even if the reported information was very simple (e.g., the mere mention that “informed consent” had been obtained). Note that we did not consider whether the provided information was truthful or not. As can be seen in Figure 3, a large proportion of studies fail to report this information.

Third, we attempted to assess whether appropriate procedures aimed at probing participants’ knowledge of the study purpose and/or hypotheses were described. We again simply coded whether each study reported having carried out such probing or not, regardless of the specific nature of the enquiry.

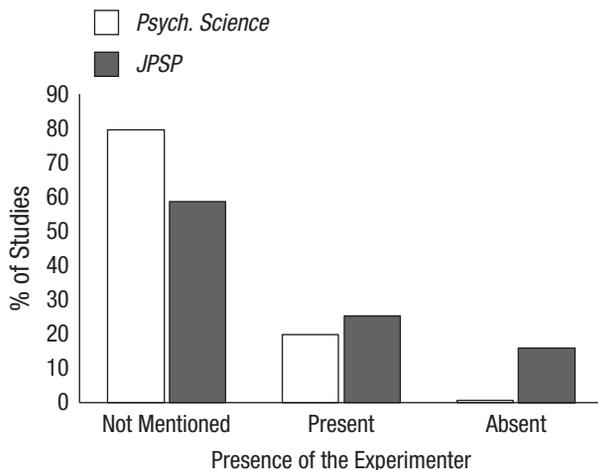


Fig. 2. Mention of the presence of the experimenter during the session.

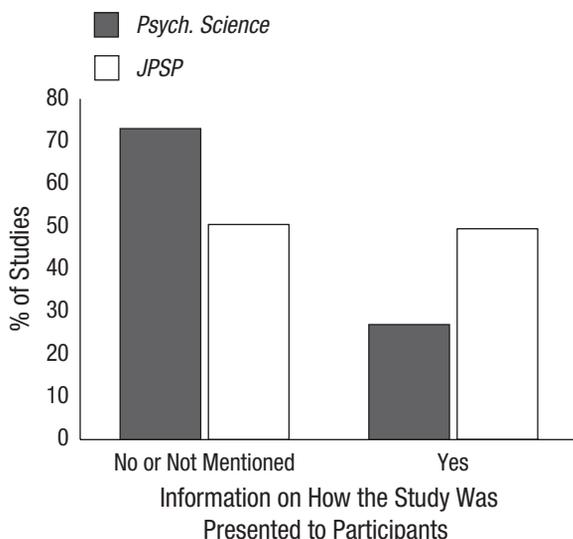


Fig. 3. Mention of information regarding the presentation of the study to participants.

As this issue is particularly relevant when participants are deceived with respect to the study’s real purpose, we also coded the presence of deception. Thus, if the researchers actively misrepresented the purpose of the study or aspects of its procedure, we considered deception to be present. For example, including a confederate posing as another participant, falsely describing computer-generated feedback as originating from a human being, or falsely presenting two parts of a single experiment as unrelated were all considered to be deception. By contrast, studies in which the real purpose of the experiment was described only vaguely or not mentioned at all were not considered to be deception. The results are presented in Figure 4. As the figure clearly shows, the majority of studies fail to report measures aimed at probing suspicion, even when deception is involved. In the absence of deceptive procedures, only a small minority of report of studies refer to awareness checks. This is a matter of concern given that the impact of demand characteristics is in no way limited to situations in which the study participant is misguided as to the real purpose of the study (Orne, 1969/2009).

Taken together, these findings indicate that, at least in the sample of articles we surveyed, relatively little attention is paid to informing the reader about the social context of the experiment. We know very little about the presence or absence of the experimenter. Similarly, only a few papers report information on participants’ expectations about the purpose of the study, even though it has become a truism that expectations guide behavior (e.g., Ajzen, 1991, Roese & Sherman, 2007). Thus, much to our surprise, it was often simply impossible to determine which precautions, such as the use of a double blind procedure (de Craen, Kaptchuk, Tijssen, & Kleijnen, 1999), had been taken to minimize experimenter bias or to gauge the impact of demand characteristics.

In view of this observation, it is worth considering the stylistic canons of a “Methods” section for a psychology experiment. The passive voice often prevails throughout (“participants were instructed to...” or “were asked to...”). As this syntactical construction does not require an actual grammatical subject, the experimental participant often appears like the object of an action performed by an unknown agent. This echoes Orne’s quote on “what is done to the subject” Academic style manuals usually recommend using the passive voice as little as possible (American Psychological Association [APA], 2010; Strunk & White, 1979; Sword, 2012), as it obscures the relationship between the subject and the action—“who is kicking whom” (Lanham, 2000; Sword, 2012). In the description of psychology experiments, the use of the active voice, far from being a mere aesthetic device, would often provide useful information regarding the social context of the experiment.

What happened to the social context of the psychology experiment?

Different reasons may account for the lack of attention to demand characteristics and expectancy effects. We identified six such factors.

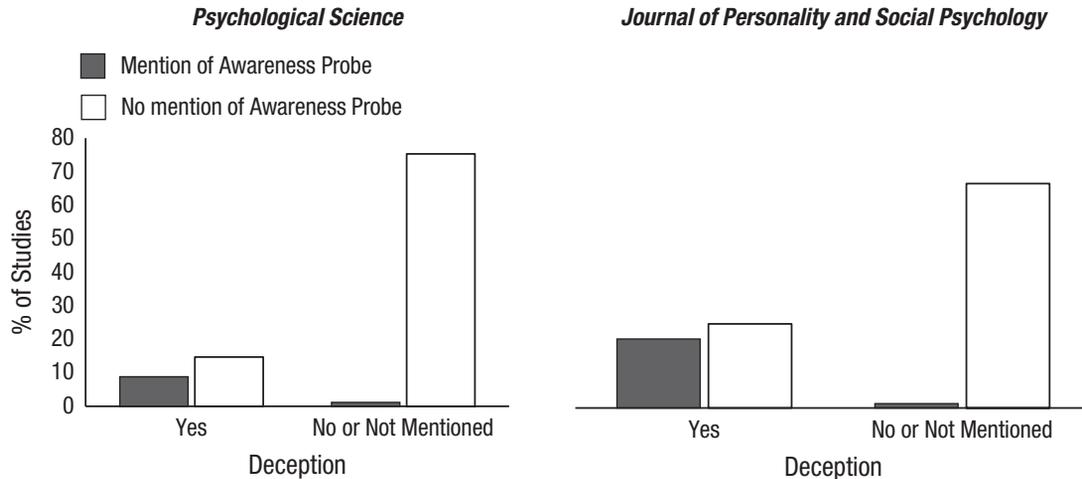


Fig. 4. Mention of awareness/suspicion probe as a function of the use of deception.

Automatization. The primary reason why contemporary psychologists may be less interested in demand characteristics and experimenter bias today lies in the automatization of experimental procedures enabled by technological developments. In the third part of his book, Rosenthal (1975/2009) put forward several methodological procedures designed to minimize experimenter bias. One of these seemed quite radical and consisted of eliminating the experimenter altogether. When the first edition of Rosenthal's book was published in 1965, social psychology experiments typically involved direct contact between the experimenter and the participant. Now, in many, if not most, psychology experiments, the participant has little direct interaction with the experimenter, who merely greets him or her upon entering the laboratory. The research participant instead most typically interacts with a computer that displays standardized instructions. Thus, with contact between the experimenter being minimized, we can expect experimenter bias to be somewhat less of a pressing issue. Rosenthal (1975/2009) anticipated this possibility:

The day may yet come when the elimination of the experimenter, in person, will be a widespread, well-accepted practice. Through the use of computers we may generate hypotheses, sample hypotheses, sample the experimental treatment conditions from a population of potential manipulations, select our subjects randomly, invite their participation, schedule them, instruct them, record and analyze their responses, and even partially interpret and report the results. (Rosenthal, 1975/2009, p. 598)

The day Rosenthal described has now arrived. Many studies are now administered via automated procedures. However, Rosenthal noted that automatization does not resolve all problems:

Even if this fantasy were reality we would not be able to eliminate the experimenter completely. He or his surrogates or colleagues must program the computer and thereby influence the output. However, there would at least be no unprogrammed differential treatment of subjects as a function of their experimental conditions. In short, although experimenter or even machine effects cannot be completely eliminated, we can at least hope for unbiased effects. (Rosenthal, 1975/2009, p. 598)

Thus, computerized procedures do not necessarily eliminate experimental demand. Participants may conform to what they regard as the experimenter's expectations as communicated through the computer. The participant may also detect the purpose of the manipulation and respond as a function of his or her understanding of its purpose rather than because of the manipulated variable per se.

Types of dependent variables and stimuli. Another explanation for the declining interest for demand characteristics and experimental demand lies in the type of dependent variables studied by contemporary researchers. When Rosenthal and Orne produced their seminal works, the dependent variables most often considered by social psychologists consisted essentially of self-report measures. Today, psychologists, inspired by cognitive science and neuroscience, rely increasingly on more sophisticated behavioral measures that may be difficult to manipulate consciously. According to this line of reasoning, the use of such measures may neutralize the effect of experimental demand. However, it seems that even methods designed to tap implicit processes may be influenced by the research participants' conscious motives (cf. e.g., Fiedler & Bluemke, 2005; Gawronski, LeBel, & Peters, 2007). If this is even merely plausible, caution with respect to demand characteristics should be in order.

Changes in psychologists' topics of interest. An additional factor that may influence the current lack of interest for demand characteristics and experimenter bias resides in the evolution of research interests: It is conceivable that the phenomena that are most popular today involve processes that are less likely to be influenced by expectancies than those that were studied in the 1960s. This would also be consistent with one of Rosenthal's early points: "There are experiments in psychophysics, learning, and psychopharmacology in which the average obtained responses may be only trivially (...) affected" (Rosenthal, 1975/2009, p. 383). It is conceivable that experiments in domains that are more immune to expectancy effects, or thought to be so, are relatively more frequent today than they were 40 years ago (although the examples provided by Rosenthal represent but a tiny fraction of the papers published in the two journals we surveyed, especially *JPSP*).

Habits. Another optimistic explanation would hold that researchers in psychology have been well-trained to avoid the pitfalls of demand characteristics and experimenter bias—so well-trained, in fact, that precautions to avoid them are well entrenched in current best practices and do not deserve to be described in any form of detail anymore.

Journal space. It is also possible that competition for journal space and the increased reliance on short reports has forced authors to reduce the method section to its bare bones, which may thus involve omitting information relevant to expectancy and demand effects. These are indeed rarely the focus of the study.

Encapsulated view of cognition. Finally, we would like to suggest that there may be a further, deeper reason why interest in experimenter bias and demand effects has declined so dramatically—namely, the fact that the rise of cognitive neuroscience has further promoted a view on cognitive processes as wholly dependent on and encapsulated in neural processes that take place within the confines of the brain (Dovidio, Pearson, & Orr, 2008; Klein, 2009). The very manner in which behavior is now routinely described as being driven by the involvement of this or that brain region eschews the fact that the brain is itself in constant interaction not only with the world, but also with other people. This focus on neural, internally driven accounts of behavior relegates social influence to the status of a mere source of noise that needs to be controlled or eliminated, much like the practice of averaging participants' performance on some behavioral task considers, perhaps erroneously so, individual differences to be no more interesting than random error (for a contrarian view, see Leach, 2010). It should be clear, however, that social influences can be just as strong as any other on performance, even with respect to phenomena thought to reflect automaticity, such as the Stroop effect (Huguet, Galvaing, Monteil, & Dumas 1999; Raz, Fan, & Posner, 2005; Raz, Kirsch, Pollard, & Nitkin-Kaner, 2006). It should in fact come as no surprise

that language, manner of speech, authority, and other such factors influence behavior, for these are factors that anyone experiences the effects of on a daily basis (e.g., anyone can profoundly change another person's emotional state by uttering a few words). Emerging fields, such as social cognitive neuroscience, have begun to acknowledge and document such effects, but it is fair to say that interest in these issues remain cursory at best insofar as mainstream cognitive neuroscience is concerned.³

Situating the Psychology Experiment

As we have already noted, Orne criticized experimental psychologists for relying on a model inspired from physics. In his view of this model, the research participant is but a mere passive receptacle of experimentally manipulated and carefully controlled stimuli. The social context in which the experiment takes place is viewed as irrelevant. Most contemporary researchers in psychology would explicitly reject this model. Our content analysis nevertheless suggests that not enough attention is being paid to the social context in which the experiment takes place. In this section, we consider recent perspectives that have questioned the prevalent view of psychological processes as context-free and encapsulated.

Cognitive psychology has long considered the workings of the human mind as subtended by encapsulated processes (e.g., Fodor's "modularity of mind") that operate on symbolic representations. According to this perspective, contextual influences are construed as "add-ons": "secondary, often paradigm-specific processes that merely moderate the mental processes of fundamental interest" (Smith & Semin, 2007, p. 134). Although this "classical" perspective on information processing (e.g., Simon, 1957) has now long been superseded, in part because of the influence of connectionism (Rumelhart, McClelland, & the PDP Research Group, 1986) and other interactionist perspectives on cognition, it nevertheless continues to permeate contemporary thinking, in part because it is a much simpler tool to think with than the alternatives.

Amongst these options, proponents of "situated cognition" (e.g., Clark, 1997) have long argued that representations emerge online as a function of contextual cues. More recently, a similar account has been applied to social cognition (Smith & Semin, 2004). According to this view, contextual cues may affect the use of concepts and representations through their impact on the perceiver's current social motives. Thus, general processes, such as person perception, may operate very differently depending on their action relevance. When people are in "TV-viewing mode" (as participants in computerized experiments often are), they may not process stimuli as they do when the very same stimuli serve a pragmatic purpose. The relevant actions associated with these stimuli are necessarily a function of the social context of the experiment: For what purpose is the participant expected to store, retrieve, or form impressions? Rather than being a form of "noise," context is integral to the appraisal of mental processes.

This observation may lead us to revise our conception of demand characteristics and experimenter bias. If we adopt this “situated cognition” approach, we need to view this social context as integral to the psychological processes under study. Observing, for example, that the experimenter’s expectations affect the outcomes of the study does not necessarily mean that the “real” processes are “polluted” by these expectations. Rather, it may tell us something about how these processes may serve a social function, such as regulating interactions with a social partner (e.g., the experimenter). Smith and Semin (2004) made the following argument:

The true strength of the laboratory is not its supposed insulation of behavior from context effects, but its flexibility in allowing experimenters to construct very different types of contexts, suited to test different types of hypotheses. But a methodological implication of this way of thinking is that laboratory researchers should be mindful of the contexts and situations they are creating. (p. 88)

To illustrate this argument, we shall consider a phenomenon that has been widely studied in social cognition: automatic behavioral priming.

A Case Study: Automatic Priming Effects

Automatic behavioral priming refers to the fact that constructs such as a personality trait or a stereotype, when activated through relevant primes, may automatically influence subsequent behavior in the absence of awareness. There are many instances of such effects in the social psychology literature (see Blair, 2002; Dijksterhuis & Bargh, 2001; Wheeler & DeMarree, 2009 for reviews). The most emblematic of such studies, and also the first one to describe the phenomenon, is Bargh, Chen, and Burrows’s (1996) study, in which they reported (in Studies 2a & 2b) that participants unwittingly walked slower after incidental exposure to words related to the concept of old age. Bargh et al.’s experiments were structured in the following manner. First, participants had to arrange words into meaningful sentences (based on the scrambled sentences task, Srull & Wyer, 1979). For half of the participants (the primed group), some of these response words were associated with the concept of old age. For the other half (the control group), all the prime words were replaced by neutral ones. Upon completion of the task, participants were dismissed. As they would exit the building through a long corridor, their walking speed was measured manually by a confederate posted along the participant’s exit path. Bargh et al. observed that primed participants walked more slowly than control participants. Debriefing showed that participants were unaware of the prime or of its effect on their behavior. Many subsequent studies using different concepts and behaviors have documented similar effects (for a review, see Bargh, 2006).

Thus far, the most prevalent explanation for this phenomenon has been the “perception–behavior link” (Dijksterhuis & Bargh, 2001), itself based on James’ *ideo-motor principle*. Thus, merely thinking about a particular behavior automatically activates the corresponding behavior and so prepares one for the relevant actions. By extension, the perception of cues associated with specific actions (e.g., seeing a door handle) likewise activate the corresponding behaviors (e.g., a grasping movement). By extension again, one may thus think that even more distant cues may likewise activate associated behaviors. Thus, for instance, an abstract concept such as “intelligence” may prime one to act in a more intelligent manner. According to this account, primed concepts automatically activate a vast array of cognitive constructs (scripts, motivations, etc.) that share the same coding as the action (see also Hommel, Musseler, Aschersleben, & Prinz, 2001) and can therefore be expressed in various behaviors. For example, seeing a backpack in an experimental room may activate “outdoors” or “wilderness” but also “cooperation” (Bargh, 2006; Kay, Wheeler, Bargh, & Ross, 2004). Early research on behavioral priming has tended to consider it as an almost unavoidable phenomenon: Primed concepts are assumed to necessarily trigger the associated behavior.

Yet, evidence suggesting that priming is susceptible to contextual influences has accumulated over recent years (Blair, 2002). For example, Cesario, Plaks, Hagiwara, Navarrete, and Higgins (2010) have found that the same prime (the stereotype of African Americans) may elicit either a fight or a flight response depending on whether participants were in a closed, confined space or in a wide open space respectively.

The perception–behavior link model acknowledges the existence of such contextual influences (cf. Bargh, 2006) and suggests that, in addition to containing the prime, the context may also “filter” the types of behavior that will actually be thus facilitated or inhibited. Further, depending on their goals, individuals will focus their attention on a specific range of potential primes or constructs. Motivation is a further important modulatory factor that may determine which specific behavior is expressed in response to a prime. Such goals and motivations are to a large extent context-driven. But in this model, contextual filtering takes place at a very late stage, selecting from among the many “streams of input” (i.e., the rich representations evoked by multiple primes) those that will influence behavior (Loersch & Payne, 2011). Thus, the prime-to-behavior link is rather complex—so much so, in fact, that Wheeler and DeMarree (2009) recently documented as many as 16 different paths through which a prime may influence (or not) behavior.

In this light, some of us (Doyen, Klein, Pichon, & Cleere-mans, 2012) recently reported a failure to replicate the walking speed study (for another unsuccessful attempt, see Pashler, Harris, & Coburn, 2011). Reflecting on why this was the case, we considered the possibility that the social context of the experiment may be responsible for this failure and thus carried

out a second study in which participants' expectations about the outcome of the study were now manipulated: Half of the experimenters were told to expect their participants to walk faster, whereas the other half were told to expect their participants to walk slower. Under these conditions, we managed to replicate the walking speed effect, but only when both primes and congruent experimenter expectancies were present.

It is possible to frame the opposition between Bargh et al.'s (1996) early findings and Doyen et al.'s (2012) in terms of internal validity. One extreme line of argument would use the Doyen study to affirm that behavioral priming, as observed by Bargh et al. (1996), reflects a form of demand effect and is merely an "artifact" due to the experimenter's expectations. In this view, the perception-behavior link model is not warranted by the method of their walking speed study. The "defense" against this attack could be that minute differences in Doyen et al.'s procedures are responsible for the failure to replicate the initial study in all conditions and/or that all precautions were made to minimize the influence of experimenter expectations and demand characteristics in the original studies (Bargh, 2012a, 2012b; Yong, 2012).^{4,5} Most notably, experimenters in Bargh et al.'s studies were reportedly blind to conditions, which renders an explanation in terms of experimenters' expectations implausible. In this view, the "noise" inherent in the variations implemented by Doyen et al. explains that they have failed to replicate Bargh's findings in their initial experiment. Note that, in this imaginary debate, both the defense and the prosecution of the original study are committed to a common assumption: Contextual variables, such as experimenters' expectations, are a source of error that obscures the process of interest. As we have suggested above, this assumption may reflect an encapsulated view of priming. But is there a more situated alternative? In the next section, we consider such an account.

Situated Perspectives on Priming

The perception-behavior link has recently been challenged by Loersch and Payne (2011), who have put forward a "situated-inference model" of automatic priming. It differs substantially from Bargh's (2006) in its emphasis on the situational determinants of priming. Indeed, this model suggests that behavioral priming occurs in three stages:

1. The primes activate related concepts and constructs, increasing their accessibility (but usually not enough to affect judgment or behavior).
2. This information is misattributed to one's natural response in the environment (e.g., what one's thinks about the task or how one feels).
3. This information is used to select a behavior depending on the "affordances" of the environment.

This last step is the most interesting for our purposes. Loersch and Payne have extended the concept of visual

"affordance" initially coined by Gibson (1979). Gibson famously suggested that visual perception was a function of physical action possibilities. Similarly, Loersch and Payne have proposed that both physical and social objects in environment afford specific questions to the individual (e.g., "How should I behave with this person?"; "What should I order at this bar?"). People use accessible information to answer these questions. In the new model, the effect of the prime will be contingent on the type of "questions" afforded by the environment. In this view, the great diversity of priming effects is not attributable to the complexity of activated representations but rather to the richness of the environment. Besides rather than filtering a process that is already under way (as suggested by Bargh, 2006), contextual influences operate at a very early stage in this model. It is worth emphasizing the parallel between this approach and the concept of demand characteristics. In Orne's view, participants consciously infer the questions addressed by the experiment and try to respond to them as adequately as possible. However, such influences may be much more subtle and implicit. As one moves toward a situated approach to psychological processes, the impact of demand characteristics may appear less and less like a source of bias and more like an integral part of the processes under study.

The situated inference model does not directly account for Doyen et al.'s findings. However, it leads us to appraise the controversy it elicited in a different light. Indeed, knowing that the experimenter's expectations influence participants' behavior may tell us something as to the type of information that the participant uses to make behavioral choices (such as walking slowly). After all, the participant leaves the lab following an injunction from the experimenter. In the perception-behavior link model, we would expect slower walking speed regardless of whether the participant leaves the laboratory following such an injunction or in response to any other contextual stimulus. If we take a more situated perspective, the presence of this injunction and the person issuing it may be crucial to the understanding of the process. It may also shape the cues used by the participant to regulate walking speed. Given that the injunction is made by the experimenter, the latter's behavior may constitute such a cue. In this view, nonverbal mimicry (Chartrand & Bargh, 1999) may for example be part of the explanation for the reduced walking speed in Doyen et al.'s replication.

More generally, Doyen et al.'s (2012) results are compatible with the existence of priming effects: It is just that the primes are presented in a social context that either also acts as a congruent prime or not. Thus, experimenters who believe their participants will walk slower may act in a manner that is congruent with these expectations (e.g., moving and speaking more slowly). Such behavior may be unconsciously picked up by participants, who would then unwittingly adapt their own behavior to that of experimenters.

Regardless of the validity of this hitherto untested account, it is worth emphasizing that it would not imply that "automatic

stereotype based behavioral priming does not exist,” but rather that it may operate very differently depending on the social context in which it takes place. More generally, this case study highlights how subtle variations in precautionary measures, which many of the surveyed papers failed to report in their method section, can dramatically change the pattern of results one observes.⁶

What Should We Do About It? Practical Recommendations

Having now reiterated the importance of the social context to the understanding of psychological processes, an obvious question arises: What should we do about it? An obvious solution, advocated in much of the work on these topics (Rosenthal, 1975/2009), involves minimizing as many sources of bias as possible. This may consist, for instance, of using double blind procedures or of avoiding contact between experimenter and participants.

A second approach (cf. Rosenthal, 1975/2009) involves introducing controls that test for demand characteristics. For example, one can ask a group of participants to respond in ways that confirm the study’s hypotheses. Orne (1962, 1969/2009) also recommended the use of quasicontrols to evaluate the influence of demand characteristics. One such quasicontrol involves presenting the study to people from the same population as the “real” participants and to ask them to imagine that they are participants themselves but without incurring the actual treatment. Participants are then asked to report their data. Another possibility involves asking participants to simulate responding as if they had been affected by the treatment (that they did not actually receive). In the same vein, several studies have now demonstrated that intentional simulations of implicit processes may actually produce consistent effects (e.g., Fiedler & Bluemke, 2005), which highlights their importance.

A third solution involves using financial incentives as a function of performance. Indeed, these make clear the demand characteristics of the experiment (Hertwig & Ortmann, 2001). This solution could also be the focus of experimental manipulations to test the effect of such characteristics even when performance is not the focal DV. For example, knowing that participants who are rewarded for conforming to the hypotheses do not exhibit the same pattern of responses to the experimental manipulations as those who do can be extremely informative.

A fourth, more widespread, approach involves measuring variables that may be implicated in the effect of demand characteristics and experimenters’ expectations. There are some specific guidelines that can be advocated in this regard.

Thorough debriefing

First, one can recommend engaging in thorough debriefing procedures (Orne, 1962). When such procedures are reported,

we usually know very little about their actual content: For example, in our study, the authors would often simply state “we used a funnel debriefing.” We believe that this information should be not only briefly reported but also described in detail as it provides an insight into participants’ expectations regarding the experiment.

In a recent paper, Froese, Gould, and Barrett (2011) described three types of approaches to consciousness. Inspired by the behaviorist tradition, experimental research in this domain generally favors third-person data (i.e., objective, public data produced by methods such as mental chronometry and neuroimaging). Froese et al. contrasted this approach with first-person and second-person approaches. The former refer to introspective data and the latter to the outcome of interviewing, as typically produced during debriefing. They suggested that such subjective data are instrumental to identifying the factors that influence performance. We should always ask the subject what he or she thinks about what is going on in the experiment, Froese et al. argued, for in doing so one may come to realize that what experimenters had intended is not what actually happens. This phenomenological approach, however, demands a thorough training of the interviewer. Froese et al. recounted the manner in which this method has allowed researchers to gain valuable theoretical insights that have led to testable, and supported, theoretical predictions (e.g., Waroquier, Marchiori, Klein, & Cleeremans, 2009, 2010). Thus, Froese et al.’s analysis compels us to give participants the central spot they should never have left in our conduct of our experimental work, for asking people how they perform a task has direct bearing on our interpretations of the data. In this sense at least, verbal reports are data, as Ericsson and Simon (1980) forcefully argued.

However, as Froese et al.’s analysis suggests, verbal reports as obtained through interviews and other such methods should be considered very cautiously. Nisbett and Wilson (1977) famously remarked that “Subjects are sometimes (a) unaware of the existence of a stimulus that importantly influenced a response, (b) unaware of the existence of the response, and (c) unaware that the stimulus has affected the response” (p. 231), further pointing out that participants will often volunteer information that is not based on fact but rather on a rational, post-hoc, reconstruction of their behavior. And indeed, suspicion probes are often inefficient even when the participant has been made clearly aware of the hypothesis (Nichols & Maner, 2008) or of the experimenter’s intention to deceive him or her prior to the experiment (Newberry, 1973). These important points, however, do not diminish in and of themselves the value of careful extended interactions with participants: One merely needs to be aware of the limits of such methods.

Integrating the social context of the experiment

Let us now come to our second guideline regarding the appraisal of demand effects. Authors should supply detailed information

regarding these three entities: the experimenters, the participants, and the experimental setup. To ensure a comprehensive reading and understanding of the experimental situation, beyond what is commonly reported, the reader should be able to find an answer to the questions highlighted in Table 1.

If double blind procedures could not be employed or were not judged desirable this should be clearly justified. Also, note that, to circumvent the problem of low journal space, this detailed material could be made available online.

If we adopt the situated cognition perspective, however, another complementary approach can be adopted. It involves designing psychology experiments in ways that make the social context of the experiment more relevant to the process under investigation. What is the purpose of the task from the participant's perspective? How do the contextual features of the experiment affect the actions fulfilled by the task? Consider, for example, a study on person perception. In such a study, knowing the purpose of the impression participants form of the target may be an essential ingredient of the study.

One could also engage in efforts to better specify to the participants the context in which the stimuli are to be evaluated. To borrow an example from Smith and Semin (2004), when evaluating Asians as part of a stereotyping study, it may be useful to specify the context of this evaluation. For example, is it in the purpose of evaluating potential business partners or competitors? Or in appraising tourists visiting one's city?

A failure to consider contextual information in the design and presentation of the tasks to study participants may lead to serious misinterpretations. In such an apparently "context-free" situation, participants may have inferred a context or

purpose and responded accordingly. But as the researchers did not specify a context and could not know what the participant inferred, it will remain unclear whether the assumed process is "general" or closely associated with this context. Obviously, this will undermine our capacity to properly replicate the experiment. In the event of a failure to replicate, it will also make it more difficult to infer the reason for it: Type II error or correct rejection of the null hypothesis?

The use of scripts, often seen in experimental economics but not in psychology (Hertwig & Ortmann, 2001), is a possible response to this problem in some branches of our discipline. Typically, scripts place participants in specific roles ("buyer" or "seller" in a negotiation game for example) and specify the behavioral options that are open to them. When combined with the presence of financial incentives, enacting these scripts does not reflect a mere form of role playing or simulation. Participants become real agents whose choices have tangible consequences for them (Hertwig & Ortmann, 2001). Note, however, that scripts can vary in their level of contextualization (e.g., being a seller of unspecified goods in an abstract market vs. a specific product in a well-known market). They may constitute a response to the problem of failed replication only insofar as they are properly situated. There is no reason to limit their use to economic settings.

Conclusion

In this article, we have revisited two old related "scarecrows" of experimental psychology: demand characteristics and experimenter bias. Fifty years after they entered the limelight, these topics have largely disappeared from the agenda of experimental psychologists. Some of the reasons for this, such as the automatization of psychology experiments, are certainly understandable. And we must acknowledge that, for certain subfields of our science (e.g., psychophysics), these phenomena pose less of a problem than for others (e.g., social cognition). Yet, when writing in 2008, Rosenthal and Rosnow (2009) noted that "problems of experimenter effects have not been solved yet" (p. ix).

We considered one natural response to demand characteristics and experimenter effects: increased methodological scrutiny and more rigorous control. Yet, we have to remain sensitive to the fact that it will never be possible to control all possible influences of these factors, for psychological experiments never take place in the ideal form that is achievable in, say, physics. Thus, participants are not zombies who will respond in a predictable manner to stimulation. Instead, they continuously exert their intelligence and attempt to guess a study's objectives, discern hard-to-see stimuli, or monitor their own actions. Both participants and experimenters also formulate conscious expectations regarding the experiment and the role they are expected to play in the study.

In this light, a central feature of conscious awareness is that it simply cannot be turned off, even when experiments are

Table 1. A List of Questions That Should Be Commonly Answered in Reports of Experimental Studies.

Experimenters	
Who are they?	
What do they know about the experiment?	
What do they know about the participants?	
What could they expect from the experiment?	
Participants	
Who are they?	
How are they rewarded?	
What do they know about the experiment?	
What do they know about the experimenter?	
What could they expect from the experiment?	
The experimental setup	
What type of reward is offered?	
Is there any form of deception?	
What information is provided to the participants?	
How is such information transmitted to the participant?	
Is the experimenter present or absent during the dependent variable measurement?	
Are the participants isolated during the experiment or are they at perceptual distance of peers?	

designed to minimize its effects. In controlled experiments, researchers attempt to minimize the effects of such conscious monitoring on people's responses so as to reduce noise and hence obtain "purer" data. However, not all such factors can be controlled, and it may be the case that both participants and experimenters remain unaware of the influence that some factors exert on the outcomes of the study. The case of Clever Hans is exemplary in this respect: Neither Van Osten (his owner) nor the horse were aware of the cues they respectively emitted and used so as to endow Hans with seemingly extraordinary mathematical skills.

In conclusion, we should like to recommend that rather than attempting to eliminate all potential sources of bias by automatizing experiments, experimental psychologists should instead strive to explore such sources of bias in a systematic manner. Failing to do so puts psychology at risk, for what sense is there in developing general conclusions about entire classes of psychological phenomena if such phenomena are tightly dependent on interactions with computers? Eliminating the experimenter and, more generally, the social context of the experiment from our understanding of psychological processes may certainly increase control and maximize internal validity, but it also exposes entire segments of psychological research to irrelevance.

Declaration of Conflicting Interests

The author declared no conflicts of interest with respect to the authorship or the publication of this article.

Notes

1. Although such demand characteristics have not always been seen negatively. On the contrary, in early experimental psychology it was considered useful that the subject be an expert, as it would increase the accuracy of responses (Despret, 2004).
2. The choice of 2005 for *Psychological Science* was dictated by the fact that, besides regular research reports, more recent volumes include very short papers with online supplementary material, which considerably complicate coding. However, we do not see any reason why the pattern observed in 2005 would have sensibly changed.
3. This "encapsulated view" of cognition, or of psychological processes in general, is actually more ancient and dates at least from behaviorism, which typically considered the relation between stimulus and response independently of the social context in which they were embedded. This affected the reporting of the studies. For example, in a typical verbal learning experiment (e.g., J. C. Cohen & Musgrave, 1966), it is very difficult to identify what the experimenter exactly told the subject, which makes replication difficult or impossible.
4. An attempt to respond to these methodological criticisms has been made through a new replication integrally available online: <http://www.replicate.vgbrain.net>
5. At the time we were writing this article, these blog posts were no longer online. Yong (2012) offers a summary.
6. Note that Bargh does not ignore demand characteristics in priming research. He has developed a detailed debriefing technique to evaluate the consciousness of the prime and its possible consequences (Bargh & Chartrand, 2000).

References

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179–211. doi:10.1016/0749-5978(91)90020-T
- American Psychological Association. (2010). *Publication manual of the American psychological association*. Washington, DC: Author.
- Bargh, J. A. (2006). What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. *European Journal of Social Psychology*, *36*, 147–168. doi:10.1002/ejsp.336
- Bargh, J. A. (2012a, March 5). *Nothing in their heads* [Web log post]. Retrieved from <http://www.psychologytoday.com/blog/the-natural-unconscious/201203/nothing-in-their-heads>
- Bargh, J. A. (2012b, March 23). *Angry birds: Nothing in their heads, Round 2* [Web log post]. Retrieved from <http://www.psychologytoday.com/blog/the-natural-unconscious/201203/angry-birds>
- Bargh, J. A., & Chartrand, T. L. (2000). The mind in the middle: A practical guide to priming and automaticity research. In H. Reis & C. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 253–285). New York, NY: Cambridge University Press.
- Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*, 230–244. doi:10.1037/0022-3514.71.2.230
- Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, *6*, 242–261. doi:10.1207/S15327957PSPR0603_8
- Cesario, J., Plaks, J. E., Hagiwara, N., Navarrete, C. D., & Higgins, E. T. (2010). The ecology of automaticity: How situational contingencies shape action semantics and social behavior. *Psychological Science*, *21*, 1311–1317. doi:10.1177/0956797610378685
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*, 893–910. doi:10.1037/0022-3514.76.6.893
- Chartrand, T. L., Maddux, W. W., & Lakin, J. L. (2005). Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 334–361). Oxford, England: Oxford University Press.
- Chase, L. J., & Chase, R. B. (1976). A statistical power analysis of applied psychological research. *Journal of Applied Psychology*, *61*, 234–237. doi:10.1037/0021-9010.61.2.234
- Clark, A. (1997). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, *65*, 145–153. doi:10.1037/h0045186
- Cohen, J. C., & Musgrave, B. (1966). Effects of formal similarity on cue selection in verbal paired-associate learning. *Journal of Experimental Psychology*, *71*, 829–838. doi:10.1037/h0023218
- de Craen, A. J., Kaptchuk, T. J., Tijssen, J. G., & Kleijnen, J. (1999). Placebos and placebo effects in medicine: Historical overview. *Journal of the Royal Society of Medicine*, *92*, 511–515.

- Despret, V. (2004). *Hans, le cheval qui savait compter* [Hans, the horse who could count]. Paris, France: Les Empêcheurs de Penser en Rond.
- Dijksterhuis, A., & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 33, pp. 1–40). New York, NY: Academic Press. doi:10.1016/S0065-2601(01)80003-4
- Dovidio, J. F., Pearson, A. R., & Orr, P. (2008). Social psychology and neuroscience: Strange bedfellows or a healthy marriage? *Group Processes & Intergroup Relations*, *11*, 247–263.
- Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind but whose mind? *PLoS ONE*, *7*, 1–7. doi:10.1371/journal.pone.0029081
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*, 215–251.
- Fiedler, K., & Bluemke, M. (2005). Faking the IAT: Aided and unaided response control on the Implicit Association Test. *Basic and Applied Social Psychology*, *27*, 307–316.
- Froese, T., Gould, C., & Barrett, A. (2011). Re-viewing from within: A commentary on first- and second-person methods in the science of consciousness. *Constructivist Foundations*, *6*, 254–269.
- Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures really tell us? scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, *2*, 181–193. doi:10.1111/j.1745-6916.2007.00036.x
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton-Mifflin.
- Harris, M. J., & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological Bulletin*, *97*, 363–386. doi:10.1037/0033-2909.97.3.363
- Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists. *Behavioral and Brain Sciences*, *24*, 383–451.
- Hommel, B., Musseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and event planning. *Behavioral and Brain Sciences*, *24*, 849–877.
- Huguet, P., Galvaing, M. P., Monteil, J. M., & Dumas, F. (1999). Social presence effects in the Stroop task: Further evidence for an attentional view of social facilitation. *Journal of Personality and Social Psychology*, *77*, 1011–1025. doi:10.1037/0022-3514.77.5.1011
- Kay, A. C., Wheeler, S. C., Bargh, J. A., & Ross, L. (2004). Material priming: The influence of mundane physical objects on situational construal and competitive behavioral choice. *Organizational Behavior and Human Decision Processes*, *95*, 83–96. doi:10.1016/j.obhdp.2004.06.003
- Klein, O. (2009). From utopia to dystopia: Levels of explanations and the politics of social psychology. *Psychologica Belgica*, *49*, 85–100.
- Klein, O., & Snyder, M. (2003). Stereotypes and behavioral confirmation: From interpersonal to intergroup perspectives. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 35, pp. 153–234). New York, NY: Academic Press. doi:10.1016/S0065-2601(03)01003-7
- Lanham, R. (2000). *Revising prose* (4th ed.). Boston, MA: Allyn & Bacon.
- Leach, C. W. (2010). The person in political emotions. *Journal of Personality*, *78*, 1827–1860.
- Loersch, C., & Payne, B. K. (2011). The situated inference model: An integrative account of the effects of primes on perception, behavior, and motivation. *Perspectives on Psychological Science*, *6*, 234–252. doi:10.1177/1745691611406921
- Newberry, B. H. (1973). Truth telling in subjects with information about experiments: Who is being deceived. *Journal of Personality and Social Psychology*, *25*, 369–374. doi:10.1037/h0034229
- Nichols, A. L., & Maner, J. K. (2008). The good-subject effect: Investigating participant demand characteristics. *Journal of General Psychology*, *132*, 151–165. doi:10.3200/GENP.135.2.151-166
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological review*, *84*, 231–259. doi:10.1037/0033-295X.84.3.231
- Orne, M. T. (1962). On the social psychology of the psychology experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*, 776–783. doi:10.1037/h0043424
- Orne, M. T. (2009). Demand characteristics and the concept of Quasi-controls. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifacts in behavioral research* (pp. 110–137). Oxford, England: Oxford University Press. (Original work published 1969).
- Pashler, H., Coburn, N., & Harris C. R. (in press). Priming of social distance? Failure to replicate effects on social and food judgments. *PLoS ONE*.
- Pashler, H., Harris, C., & Coburn, N. (2011, September 15). *Elderly-related words prime slow walking*. Retrieved from <http://www.PsychFileDrawer.org/replication.php?attempt=MTU%3D>
- Pfungst, O. (1965). *Clever Hans (The horse of Mr. Von Osten)*. New York, NY: Holt, Rinehart and Winston. (Original work published 1911).
- Raz, A., Fan, J., & Posner, M. (2005). Hypnotic suggestion reduces conflict in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 9978–9983. doi:10.1073/pnas.0503064102
- Raz, A., Kirsch, I., Pollard, J., & Nitkin-Kaner, Y. (2006). Suggestion reduces the Stroop effect. *Psychological Science*, *17*, 91–94. doi:10.1111/j.1467-9280.2006.01669.x
- Roese, N. J., & Sherman, J. W. (2007). Expectancy. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (pp. 91–115). New York, NY: Guilford Press.
- Rosenthal, R. (2009). Experimenter effects in behavioral research. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifacts in behavioral research* (2nd ed., pp. 287–666). Oxford, England: Oxford University Press. (Original work published 1975).
- Rosenthal, R. (2009). Interpersonal expectations: Effects of the experimenter's hypothesis. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifacts in behavioral research* (pp. 138–210). Oxford, England: Oxford University Press. (Original work published 1969)
- Rosenthal, R., & Rosnow, R. L. (2009). Preface. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifacts in behavioral research* (pp. 3–5).

- Oxford, England: Oxford University Press. (Original work published 1969)
- Rosenthal, R., & Rosnow, R. L. (2009). A preface to three prefaces. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifacts in behavioral research* (pp. IX–XI). Oxford, England: Oxford University Press.
- Rumelhart, D. E., & McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing*. Cambridge, MA: MIT Press.
- Simon, H. A. (1957). *Models of man: Social and rational*. Oxford, England: Wiley.
- Smith, E. R., & Semin, G. R. (2004). Socially situated cognition: Cognition in its social context. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 36, pp. 53–117). New York, NY: Academic Press.
- Smith, E. R., & Semin, G. R. (2007). Situated social cognition. *Current Directions in Psychological Science*, 16, 132–135. doi:10.1111/j.1467-8721.2007.00490.x
- Srull, T. K., & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology*, 37, 1660–1672. doi:10.1037/0022-3514.37.10.1660
- Strunk, W., & White, E. B. (1979). *The elements of style* (3rd ed.). New York, NY: Macmillan.
- Sword, H. (2012). *Academic writing style*. Cambridge, MA: Harvard University Press.
- Waroquier, L., Marchiori, D., Klein, O., & Cleeremans, A. (2009). Methodological pitfalls of the unconscious thought paradigm. *Judgment and Decision Making*, 4, 601–610.
- Waroquier, L., Marchiori, D., Klein, O., & Cleeremans, A. (2010). Is it better to think unconsciously or to trust your first impression? A reassessment of unconscious thought theory. *Personality and Social Psychological Science*, 1, 111–118. doi:10.1177/1948550609356597
- Weber, S. J., & Cook, T. D. (1972). Subject effects in laboratory research: An examination of subject roles, demand characteristics, and valid inference. *Psychological Bulletin*, 77, 273–295. doi:10.1037/h0032351
- Wheeler, S. C., & DeMarree, K. G. (2009). Multiple mechanisms of prime-to-behaviour effect. *Social & Personality Psychology Compass*, 3, 566–581.
- Yong, E. (2012, March 10). *A failed replication draws a scathing attack from a psychology professor* [Web log post] Retrieved from <http://blogs.discovermagazine.com/notrocketscience/2012/03/10/failed-replication-bargh-psychology-study-doyen/>
- Young, E. (2012). Replication studies: Bad copy. *Nature*, 485, 298–300. doi:10.1038/485298a