

The Red Herring technique: a methodological response to the problem of demand characteristics

Cara Laney · Suzanne O. Kaasa · Erin K. Morris ·
Shari R. Berkowitz · Daniel M. Bernstein ·
Elizabeth F. Loftus

Received: 12 April 2006 / Accepted: 23 June 2007
© Springer-Verlag 2007

Abstract In past research, we planted false memories for food related childhood events using a simple false feedback procedure. Some critics have worried that our findings may be due to demand characteristics. In the present studies, we developed a novel procedure designed to reduce the influence of demand characteristics by providing an alternate magnet for subjects' natural suspicions. We used two separate levels of deception. In addition to giving subjects a typical untrue rationale for the study (i.e., normal deceptive cover story), we built in strong indicators (the "Red Herring") that the study actually had *another* purpose. Later, we told subjects that we had deceived them, and asked what they believed the "real purpose" of the study was. We also interviewed a subset of subjects in depth in order to analyze their subjective experiences of the procedure and any relevant demand. Our Red Herring successfully tricked subjects, and left little worry that our false memory results were due to demand. This "double cross" technique may have widespread uses in psychological research that hopes to conceal its real hypotheses from experimental subjects.

Pick up any introductory psychology textbook and you will find a discussion of "demand characteristics". Psychological researchers of every stripe worry that subjects will invalidate research findings by trying to behave like "good" subjects. Experimental volunteers want to do well, the argument goes, so they search for clues as to what the research is really about. The clues, in effect, cause subjects to feel that a certain kind of response is demanded—thus the term *demand characteristics*. Many subjects act on this demand and comply with what they believe researchers want from them. Other subjects may feel unduly pressured by these experimental constraints or, alternately, defiant of them, and thus intentionally behave in a way that will not support the hypotheses of the experimenters. Either way, demand characteristics can make experimental results hard to interpret or actually meaningless. They can also make manuscript reviewers wary of studies that use deception.

We begin by outlining the problem of demand characteristics in psychological research in general and reviewing how this problem has been addressed in one specific area of psychological research, namely memory distortion. We then present a new proposed method, the Red Herring technique, designed to reduce the risk of demand characteristics. Finally, we discuss how semi-structured interviews after an experiment may provide insight into the classification and analysis of responses from subjects who are particularly at risk of succumbing to demand.

Demand characteristics

The concept of demand characteristics was brought to the forefront of psychology by Orne (1962), who described how cues from the research context might inadvertently influence subjects who divine the true purpose of the

C. Laney · S. O. Kaasa · E. K. Morris · S. R. Berkowitz ·
E. F. Loftus
University of California, Irvine, CA, USA

D. M. Bernstein
University of Washington, Seattle, WA, USA

D. M. Bernstein
Kwantlen University College, Surrey, BC, Canada

C. Laney (✉)
Forensic Psychology, University of Leicester,
106 New Walk, Leicester LE1 7EA, UK
e-mail: cl136@le.ac.uk

experiment, leading them to modify their behavior to match these expectations. This concern stems from the fact that the roles of experimenter and subject are well understood and a mental script exists for the experimental interaction. Subjects knowingly and willingly place themselves under the control of the experimenter and agree to do almost any task they are asked to perform (no matter how strange it may seem). Although the individual tasks may appear meaningless (due to the design of the experiment, where the researchers attempt to hide the meaning of the experiment from the subject in order to avoid biasing effects), the mere fact that a subject is in an experiment adds purpose to the situation. Generally, individuals are identified as being “at risk” for demand behavior if they show an accurate understanding of the true purpose of the study.

Beyond the bounds of the introductory textbook, demand characteristics are a topic of current research in experimental psychology. A quick search of *PsycINFO* using “demand characteristics” as a keyword garnered 545 hits, nearly 100 of which represented peer-reviewed articles from the last decade alone. Papers addressing demand characteristics came from all corners of psychology, and conclusions were varied.

Demand in memory distortion studies

The problem of demand characteristics has been a subject of debate in the memory distortion literature. Some studies suggest that demand characteristics should not be ruled out as an explanation for experimental findings, while others find no evidence for demand characteristic effects in their results. Orne’s (1959, 1979) “real-simulating” paradigm has provided some evidence for demand behavior. This approach involves asking subjects who test low on hypnotizability to pretend that they are being hypnotized. Subjects are told the experimenter will not know if they are faking or not, and that they should attempt to avoid detection. The way to test for demand characteristics in this setting is to compare the results of the real (high hypnotizable) subjects with the simulators. If their responses are similar, then demand characteristics cannot be ruled out. Experiments that have used Orne’s real-simulating paradigm have found evidence mostly supporting the demand characteristic explanation (e.g., Bryant, Barnier, Mallard, & Tibbits, 1999), although some results do not support the expected demand conclusion (Bryant & Mallard, 2005).

Other approaches to studying the problem of demand in memory distortion research find little evidence for demand as an explanatory factor. Studies that have created conditions of high and low-demand for subjects in memory studies have failed to find a link between high demand and increased memory distortion (Bjorklund et al., 2000;

Weinberg, Wadsworth, & Baron, 1983). Other researchers have given subjects who might be inclined to produce a demand response an incentive to stop doing so (Lampinen, Neushatz, & Payne, 1999; Lindsay, 1990; Macleod, 1999). For example, MacLeod (1999) conducted a directed forgetting study where subjects were offered money to recall words they had previously been instructed to forget. Few subjects were able to recall these words, in spite of the monetary incentive, suggesting that genuine forgetting had occurred.

All of these approaches share a common quantitative experimental approach. In the first study presented here, we offer a new technique for assessing and controlling for the problem of demand in deceptive experimental studies. This new technique makes it more difficult for subjects to determine the true goals of a deceptive study (which in turn makes it more difficult for them to be subject to, or defiant of, demand), and also makes it easier for researchers to partial out those subjects who are more likely to be subject to demand, so that their results can be analyzed separately from the those of the bulk of subjects.

Various qualitative techniques have also been employed to assess demand in research studies. These techniques focus on the thoughts and intentions of subjects rather than their outward behaviors (see Loftus, Miller, & Burns, 1978, Experiment 2). By focusing on these subjective interpretations, researchers can potentially assess whether subjects have guessed the hypothesis of the experiment and then made a conscious decision to mold their behavior to fit this outcome.

Simply asking subjects about their thought processes lacks the experimental rigor of controlled conditions and manipulations, but it does allow researchers to assess directly the issue of subject motivation and decision-making in the experiment. Qualitative questions and interviews can add information from a new source and provide in-depth responses that are impossible with subtler cover-story approaches to demand characteristics. In fact, Orne (1962) proposed using the post-experimental interview as a method for identifying demand characteristic subjects. These interviews allow for in-depth questioning that moves from general introductory questions to more specific questions, which give the subject a better opportunity for expressing suspicions about the experiment and explanations for their behavior (Page & Scheidt, 1971).

In the current research (Study 2), we use a combination of experimental data on response behavior and interview data about subjects’ internal reactions and thought processes. In this way we may compare what subjects were thinking to their actual responses, in order to measure what, if any, effect suspicion and other social demand influences may have on memory distortion results.

The Red Herring technique

Many researchers try to minimize the demand characteristic problem by giving subjects a false rationale for the study. In our research on false memories, for example, we have told subjects that the purpose of the study is “to look for relationships between people’s personalities and their food preferences” when actually the purpose was to plant false beliefs and memories and explore the processes by which these take hold. Occasionally subjects will figure out what the real purpose of the study is, and may then alter their own behavior accordingly.

In response to our worry that demand may be playing a role in our findings, we devised a new method for minimizing demand characteristics that we call the Red Herring technique. If normal deception (including providing subjects with a false rationale for the study) is a cross, then the Red Herring technique is a double cross. The Red Herring technique allows naturally curious subjects to “figure out” what the study is about without *actually* figuring out what the study is about (and thus becoming subject to demand). Although we have successfully used it in our false memory procedure, it is applicable to a wide range of studies in psychology, especially those involving deception. Since the technique will be described in the context of a false memory study, we first present some background on the relevant field.

In the last decade or so, investigators have successfully planted wholly false memories into the minds of subjects (e.g., Garry, Manning, Loftus, & Sherman, 1996; Hyman, Husband, & Billings, 1995; Loftus & Pickrell, 1995; Wade, Garry, Read, & Lindsay, 2002). In these studies, subjects are told that the experimenters have special information about events that have happened to them—this information may be attributed to subjects’ parents, to historical records, or to a clever computer program that has analyzed the subjects’ responses to questionnaires—and then asked to describe their memories of these events. In our own lab, we have given subjects false childhood memories of getting sick after eating specific foods (e.g., Bernstein, Laney, Morris, & Loftus, 2005a; b), among other events. These studies, using a very simple false feedback procedure, have produced false beliefs or memories in a sizable minority of subjects, sometimes exceeding 40%.

Some critics (including reviewers of our previous manuscripts) have suggested that “false memory” results may be due in part to the influence of demand characteristics. We have gone to great lengths in our previous studies to limit the impact of demand characteristics by providing subjects with plausible false rationales for our studies (complete with extra questionnaires designed to give this explanation credibility), by imbedding all critical items in long and complicated lists of related items, and by varying the formatting of questionnaires that are given both pre- and post-test. That is

to say, we do our best to deceive subjects, just as researchers in many areas of psychology have done for decades. But is this enough? Occasionally subjects have made unsolicited claims on our questionnaires, or when informally questioned after a study, that they believe that we are trying to implant false memories. But more subjects may deduce this, even if they do not go to the trouble of telling us so. Thus, finding a method that will minimize the likelihood that subjects will deduce our real purpose is a worthy goal.

We developed a new technique designed to simultaneously assess and limit the impact of demand characteristics. In the current Study 1, we gave subjects our standard untrue rationale (cover story) for the study—that it was “designed to look for relationships between people’s personalities and their food preferences”. The true goal of the study was to implant false food memories into the minds of our subjects and assess the consequences of those false memories (e.g., willingness to eat the food). But, crucially, there was another aspect to the study: the Red Herring. The Red Herring is an extra layer of information (separate from both what subjects were told and what we were actually studying), intended to provide a plausible explanation for the tasks subjects are asked to complete. That is, we are *doubly* deceiving subjects. In the present Study 1, the Red Herring was the problem of obesity in American children. That is, we gave subjects the cover story, but they were led to believe, through strategically placed questions and recurring themes in questionnaires, that we were *really* studying childhood obesity. One of the study’s questionnaires was entirely about subjects’ knowledge of this problem and their feelings about it. Subjects were never told outright that we were studying childhood obesity. Instead, it was an underlying theme that they could feel smart for recognizing. At the end of the study, we told subjects that we had deceived them, and then asked (through both open-ended and multiple choice measures) what they believed the “real purpose” of the study was.

The function of the Red Herring technique is to provide a specific magnet (besides the true purpose of the study) for subjects’ suspicions that they are being deceived. Because an experimental subject is a natural problem-solver who “sees it as his task to ascertain the true purpose of the experiment and respond in a manner which will support the hypotheses being tested” (Orne, 1962, p. 779), we provided subjects with a potential solution to the problem that was unrelated to the actual goals of the study. That way, even if subjects did their best to behave like “good” subjects and deliberately provided the answers that they thought we wanted, they ultimately failed because they were providing deliberate answers to the harmless questions that tapped into the Red Herring. An analogous technique was used to great effect by Agatha Christie. She long ago mastered the practice of planting Red Herrings as clues to divert her readers away from the true perpetrator of the crime.

In addition to presenting the novel Red Herring technique, we also examine the effects of demand on a set of false memory results. Although our Red Herring technique successfully tricked the majority of subjects about the purpose of the false memory experiment, a small minority did become correctly suspicious about the specific nature of our manipulation. Because these subjects could be identified by our study procedures, we could analyze their data separately from those of other subjects in order to discover whether knowledge about the true goals of the study affected subject responses in our false memory study.

A subset of the sample from Study 1 was subsequently interviewed about their experiences in the study and perceptions of it (Study 2). This one-on-one interview format allowed for the analysis of a whole different range of questions: Do suspicious subjects react differently than non-suspicious subjects to our false memory manipulation? Are there other subjects who might also be “at risk” for developing demand characteristics in this type of false memory experiment? Study 2 addresses these issues both quantitatively and qualitatively.

Study 1

Briefly, in Study 1, we attempted to plant false childhood memories about an experience with a particular food, asparagus. The true purpose of the study was to plant a false belief that subjects loved asparagus the first time they tried it as children, and to then explore whether this would make them want to eat asparagus more today. For other subjects we planted the belief that they got sick eating asparagus as children and explored whether they would avoid eating it. As a cover story, we told our subjects that the study was designed to analyze the relationship between food preferences and personality. But we also planted numerous clues designed to lure them into believing that it had a different purpose. This was the Red Herring—a second plot, if you will, leading the subjects to surmise that the study was about the problem of obesity in America. If the Red Herring worked as hypothesized, when subjects were questioned at the end of the study about their perceptions of its purpose, many should have chosen the Red Herring, and few should be left showing awareness of the study’s true purpose.

Method

Subjects

The subjects were 187 undergraduate students from the University of California, Irvine who participated for course credit. These subjects were primarily female (71%), and had a mean age of 20.2 (SD = 3.6).

Subjects were randomly assigned to four separate groups before they arrived in the lab. One group of subjects (the “Sick” group; $n = 63$) received the suggestion that they had become sick after eating asparagus as young children. Another group (the “Love” group; $n = 63$) was told that they had loved asparagus the first time they tried it as children. Finally, there were control groups for each of these experimental groups (with n s of 30 and 31, respectively).

Materials and procedure

Subjects came to our laboratory for two sessions, approximately one week apart. In the first session, they were told that they were participating in a study “designed to look for relationships between people’s personality and their food preferences”. They then completed six short questionnaires that did not conflict with this assertion. These questionnaires included two individual difference measures, questionnaires that assessed subjects’ recent experiences with various foods and food-related activities (including both asparagus and dieting) and their preferences for various foods, and a food history inventory (FHI). The FHI contained 19 items, including one of two critical items, “Got sick after eating asparagus” (for the Sick group and the Sick control group) or “Loved asparagus the first time you tried it”, (for the Love group and the Love control group), in the 11th position. The relevant critical item was embedded in a list that included other food-related items such as “Were forced to go on a diet” and “Ate freshly picked vegetables”. Subjects were instructed to rate each item on a scale anchored at 1 = *definitely did not happen* and 7 = *definitely did happen* before the age of 10.

Subjects were told that we would enter their data into our computer and provide personalized feedback when they returned the next week. Subjects were in fact given feedback profiles when they returned, but this feedback was far from personalized. Instead, there were just three versions of the profile. In all versions, subjects were told, on a single page with their names printed at the top and specific items listed in boxes so that they would look like computer-generated forms, that as young children they had disliked spinach, had enjoyed eating pizza, and felt happy when a classmate brought sweets to school. These events were designed to ring true for most subjects and thus give credibility to the critical false item. Control subjects’ profiles had just these three items. Experimental subjects received one additional (critical false) item, in the third position. They were told either that (as a young child) “you got sick after eating cooked asparagus”, or “you loved to eat cooked asparagus”.

After reading their profiles, subjects answered a question designed to ensure that they actually read and processed the profile items: “What is the most important childhood, food-related event in your life that the profile did not report”?

Then, they completed seven additional questionnaires. The first was a variation on the FHI completed on their first day in the lab (with different formatting and a scale ranging from “-3” to “3”). The second, third, and fourth questionnaires assessed subjects’ willingness to eat and pay for specific foods and their preferences for specific foods (i.e., false memory consequence measures). Each of these questionnaires included both asparagus items and items that fit with the obesity Red Herring (e.g., particularly unhealthy foods). The fifth questionnaire asked subjects to define their experiences of each of three events, including their critical FHI item (“you loved asparagus the first time you tried it” or “you got sick after eating asparagus”) and a Red Herring relevant item, “you were overweight for a significant portion of your childhood”, as a specific memory, a belief (but not a specific memory), or neither. The sixth questionnaire assessed subjects’ knowledge of and feelings about the problem of obesity in the United States.

The final questionnaire (our “Demand Check”) informed subjects that

Many psychology studies use deception. The purpose of deception is to keep subjects from responding a particular way because they know what the experimenters are looking for. This study used deception to hide the true nature of the study. That is, we told you this is a study about Food Preferences and Personality, but this is not what the study is really about. By this point in the study, you may have your own theory about what the study was *really* about. Please use the space below to provide your best guess.

The text was worded this way in order to allow subjects to feel clever for providing an answer other than what they had been told the study was about (and as such to thoroughly counteract any demand that this question itself may have presented). Subjects gave open-ended responses to this question. Then, on the following page, subjects were asked to respond to the item: “I believe the study was really looking at ...” Subjects were given nine possible goals of the study (including both the truth and the Red Herring), plus a “none of the above” option (for the full text of all nine possible study goals, see Table 1). They were asked to select the response that best described their view of the purpose of the study. Subjects were not permitted to flip back to their open-ended responses after seeing the multiple-choice question.

Results

Red Herring

Subjects’ responses to the open-ended Demand Check item (in which they were told that they had been deceived and

were asked to guess the true nature of the study) were coded into four categories: false memory, memory or response consistency, Red Herring, or totally incorrect. Responses were coded as “false memory” if they demonstrated a basic understanding of the true nature of the study. Subjects were not required to use the words “false memory”. For example, one subject wrote¹, “I think that this study was about memory and if you can convince someone of a certain memory that was not there to begin with”.

Responses were coded as “memory or response consistency” if subjects seemed to know that the study had something to do with memory, but did not demonstrate any knowledge of false memory or the expectation of response change over the course of the study. For example, one subject wrote, “The theory most likely refers to childhood memory”. Responses were coded as “Red Herring” if they mentioned obesity, eating disorders or healthy eating. Typical responses included, “My guess is that you guys are researching the prevalence of eating disorders among college students”, and “I would guess that this is a study about obesity”. Responses that could not be coded in any of these ways made up the final category. For example, one subject wrote, “I think the study was trying to relate food preference with the brain, to see which type of people who like certain foods use a different part of the brain area”.

Importantly, responses were coded as conservatively as possible, such that any response that could possibly be understood as describing false memories was coded as “false memory”. Also, earlier categories superseded later categories, such that if a subject mentioned both false memory and obesity, his or her response was coded as “false memory”. For example, the response “to see our memory recall ... also to test our knowledge of obesity and see how viewpoints on food and our relationships with food” was coded as “memory or response consistency”.

Very few subjects gave responses that referred to the original cover story for the study, presumably because the question itself informed them that this would be an incorrect answer. Two coders independently rated the responses of each subject. Complete agreement, across the four categories, was 86%. Disagreements were resolved through discussion.

These specific categories were used because they best address the issue of demand characteristics. Subjects who believed they were participating in a study of false memory would be subject to the type of demand characteristics that could influence our results, potentially making it look like we were implanting false memories when we were not in fact doing so. Those who believed they were participating in a study of memory *consistency*, by contrast, might try to

¹ All participant responses are direct quotations.

Table 1 Crosstabulations of open-ended demand check response and multiple choice responses to the item, “I believe the study was really looking at ...” in Study 1

Open-ended response category	Demand check multiple choice item					Total	
	False memory	Memory or response consistency	Red Herring	Other incorrect	<i>n</i>	%	
The relationship between personality differences and food preferences ^a	0	0	0	5	5	2.7	
The problem of obesity in American children ^b	1	4	52	6	63	33.7	
The specificity of college students' childhood food memories	0	12	7	7	26	13.9	
How college students respond to questions about their eating habits	0	10	6	3	19	10.2	
Whether there are consequences of false memories ^c	12	17	4	5	38	20.3	
The relationship between eating habits and family structure	0	4	7	2	13	7.0	
Other choices ^d	0	3	0	1	4	2.1	
None of these	1	5	2	4	12	6.4	
Multiple answers (not including “consequences of false memories” answer)	0	2	4	1	7	3.7	
Total	14	57	82	34	187		

Open-ended item was completed first

^a The official explanation of the study that subjects were given (which they were told was untrue as part of the directions for the open ended item on the preceding page)

^b The Red Herring

^c The true purpose of the study

^d Made up of: “The relationship between liking of spinach and consumption of cheesecake”, “Whether ethnicity affects how college students tend to use scales”, and “Whether people who tend to space out also tend to have masculine characteristics”

be consistent from one occurrence of a question to the next, but this outcome (if subjects were successful at it) would be counter to our hypotheses, and thus make our conclusions all the more conservative. Likewise, those subjects who believed they were participating in a study of American obesity (or, equivalently, of healthy eating) might feel the need to claim that they eat fewer donuts than they actually do, but they would not feel pressured to claim that asparagus had once made them sick, or that they had loved it the first time they tried it.

The responses of 82 subjects (44%) fit best in the Red Herring category, which was more popular than any other category. Just 14 responses (8%) were best labeled “false memory”, 57 responses (31%) were categorized as “memory or consistency”, and the remaining 34 responses (18%) could not be categorized in any of these ways. Thirty-nine subjects (21%) specifically used the word “obesity” in their open-ended responses. Of these, 35 responses were categorized as Red Herring (because there was no mention of memory). There were no significant group differences in these responses, $\chi^2(9; n = 187) = 8.83, P = 0.45$. That is, experimental condition did not predict which open-ended responses subjects would give. In particular, those subjects in the experimental conditions (whom we were attempting

to give false memories) were no more likely than controls to say that they were in a false memory study, $\chi^2(1; n = 187) = 0.86, P = 0.35$.

The multiple-choice data paralleled the open-ended data (see Table 1). On the multiple choice section of the Demand Check questionnaire, the most popular response to the prompt “I believe the study was really looking at ...” was “the problem of obesity in American children”, our Red Herring, chosen by 63 subjects (34%). Far fewer subjects ($n = 38$; 20%) chose the real purpose of the study, “whether there are consequences of false memories”. The remaining 86 subjects (46%) chose one (or more) of the other seven possible responses, or “none of the above”. Only five subjects (3%) chose “the relationship between personality differences and food preferences” (that is, the original justification subjects were given for the study), presumably because the directions on the previous page (for the open-ended item) specifically told them that this was incorrect.

There were, however, some group differences in multiple-choice responses, $\chi^2(27; n = 187) = 44.31, P = 0.02$. In particular, experimental group subjects (collapsed across Love and Sick) were more likely than controls to say that they were participating in a false memory study, $\chi^2(1; n = 187) = 6.15, P = 0.01$. We note that for many of

these subjects, the correct explanation (false memory) had to be staring them in the face before they recognized it. In particular, of the 38 subjects who chose “false memory” on the multiple-choice item, 26 of them had failed to give an open-ended response that could be coded as “false memory” (an additional four subjects gave “false memory” open-ended answers but then selected another answer on the multiple-choice item).

False memories and demand

As it is evident in the left side of Fig. 1, Love group subjects’ confidence that they had loved asparagus the first time they tried it increased from pre- to post-manipulation, but that of the Love controls did not. Statistically speaking, a repeated measures ANOVA comparing pre- to post-manipulation confidence of Love and Love control groups gave a significant interaction of time point and group, $F(1,92) = 14.27$, $P < 0.001$. We note that confidence levels for these two groups were not significantly different before the manipulation, $t(92) = 1.72$, $P = 0.09$. Subjects in both the Sick and Sick control groups became significantly more confident that they had become sick after eating asparagus from pre- to post-manipulation, such that the repeated measures ANOVA performed on these two groups yielded a significant main effect of condition, $F(1,91) = 6.35$, $P = 0.01$, but not a significant interaction, $F(1,91) = 1.51$, $P = 0.22$.²

A total of 40 subjects gave “false memory” responses to at least one of the two Demand Check items (20 of these subjects were in the Love group, 14 were in the Sick group, and six were controls). When we limited our analyses to the remaining 147 subjects, our overall results changed very little. These data can be seen on the right side of Fig. 1. A new repeated measures ANOVA comparing pre- to post-manipulation confidence in the Love and Love control groups again gave a significant interaction of time point and group, $F(1,68) = 9.81$, $P = 0.006$. (Once again the two groups were not significantly different before the manipulation, $t(48.8) = 1.50$, $P = 0.14$).³ And a new repeated measures ANOVA for the Sick and Sick control groups again failed to provide a significant interaction effect, $F(1,75) = 2.27$, $P = 0.19$. (Once again the two groups were significantly different before the manipulation, $t(57.1) = 2.05$, $P = 0.045$).

² We cannot say why the control condition increased significantly, except that due to an error of random assignment the two groups were not equivalent before the manipulation, $t(76.0) = 2.15$, $P = 0.03$, and there seems to have been a floor effect prior to the manipulation for the control group (such that their confidence had nowhere to go but up). Adjustments have been made to the degrees of freedom because of unequal variance in this and other calculations throughout this paper.

³ Adjustments have been made to the degrees of freedom because of unequal variance in this and other calculations throughout this paper.

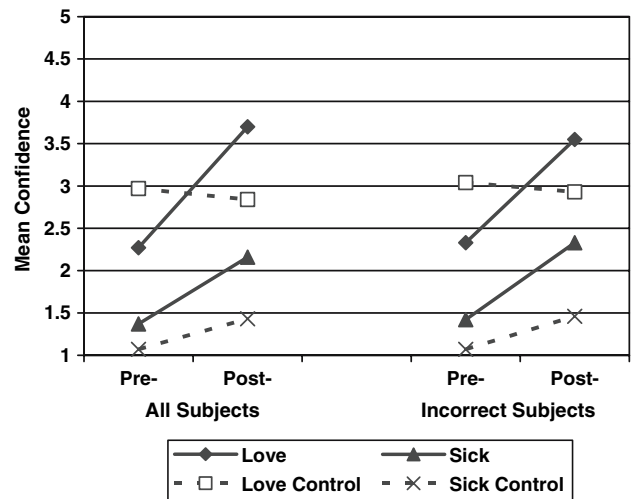


Fig. 1 Mean confidence that one of two critical false events (loving asparagus or getting sick on it as a child) had occurred, pre- and post-manipulation, for all subjects ($n = 187$), and for those subjects *not* providing “false memory” responses on either the open-ended or multiple-choice Demand Check items (“incorrect subjects”; $n = 147$). The full Love and Sick groups each had 63 subjects; the Love control and Sick control groups had 31 and 30, respectively

We also looked separately at just those 52 subjects who gave “red herring” responses on both the open-ended and multiple-choice Demand Check items. Once again, a repeated measures ANOVA produced a significant interaction of time point and group for the Love ($n = 19$) and Love control ($n = 6$) groups, $F(1,23) = 5.15$, $P = 0.03$, but not for the Sick ($n = 16$) and Sick control ($n = 10$) groups, $F(1,24) = 0.01$, $P = 0.94$.

Approximately 40% ($n = 25$) of the subjects in the Love group ended the study with false beliefs about loving asparagus the first time they tried it, according to our conservative definition of “believers” (that is, people who came to believe the suggested items were true of them; see Morris, Laney, Bernstein, and Loftus (2006) for a detailed discussion of these concepts). Put another way, our manipulation was successful for 40% of Love group subjects. Fewer subjects in the Sick group ($n = 7$; 11%) developed false beliefs of getting sick after eating asparagus. When the 40 “false memory” responders were removed from these calculations, these proportions changed very little. Of the (now smaller) Love group, 17 subjects (40%) developed false memories, whereas of the Sick group, six subjects (12%) developed false memories (see Table 2). When only “red herring” responders were included, nine Love group subjects (47%) and 1 Sick group subject (6%) developed false memories.

Statistically speaking, those who gave “false memory” responses (to either open-ended or multiple choice Demand Check items) were no more likely to become believers than those who did not give “false memory” responses

Table 2 Number of subjects meeting criteria for “believers”, by experimental condition and demand check response in Study 1

Believers, from among:	Experimental condition			Total
	Love group	Sick group	Controls	
	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	
All subjects	25 (40)	7 (11)	3 (5)	187
Subjects not providing “false memory” response(s)	17 (40)	6 (12)	3 (5)	147
Subjects providing “false memory” response(s)	8 (40)	1 (7)	0 (0)	40
Subjects providing Red Herring response(s)	14 (40)	3 (10)	2 (7)	93

“False memory” and “Red Herring” responses may have been to either open-ended or multiple-choice Demand Check items. “Believers” are those subjects who came to believe the suggested items were true of them and subsequently formed false memories

χ^2 (1; $n = 187$) = 0.36, $P = 0.55$, or than those who gave “red herring” responses χ^2 (1; $n = 124$) = 0.36, $P = 0.84$ (four subjects were excluded from this analysis because they fell into both categories).

Discussion

A major finding of the present study is that the Red Herring lured more than a third of subjects to falsely believe that they were participating in a study of childhood obesity, even though they were never told anything about this directly. Many fewer subjects deduced the actual goals of the study. Because of our use of the Red Herring methodology, we were able to identify a sizable group of subjects who were potentially susceptible to demand characteristics (because they deduced, at some point during their participation, the true goals of the study) and analyze them separately from other subjects.

We are not arguing that the 40 “false memory” responders in the present study all determined the exact goals of the study early on in the procedures and then adjusted their behavior accordingly (either to support our hypotheses or to thwart them). In fact, individual open-ended responses to the Demand Check questionnaire suggest that some subjects were not suspicious of our goals until they read the text of the open-ended question and learned that they had been deceived, and then correctly guessed the goals of the study. For example, one subject’s “false memory” response was,

I really have no idea at all because up until this point I thought it was food preferences and personality. At least with all of the questions and everything that seemed to be the direction it was going. I would say the study will have something to do with memory and how we pull memories from early childhood. It must have something to do with reconstructive memory and how certain ideas may be planted falsely into our minds. Still other subjects ($n = 26$) figured out that

they were in a study of false memory only when that possibility (in the form of an item on our multiple-choice demand check) was staring them in the face. Rather, we argue that the possibility of demand cannot be excluded for these subjects, and that we want to be conservative in addressing possible contamination of data. Therefore, these data should be analyzed separately from those of other subjects, and any discrepancies should be addressed before we can draw broad conclusions about the full dataset.

By separately analyzing the data from our “false memory” responders, we learned that demand characteristics were *not* a major factor in our false memory effect. That is, subjects who correctly deduced that they were participating in a study of false memories were no more likely to produce evidence of false memories than subjects who completely failed to deduce the true nature of the study.

The Red Herring technique is designed to make this process of identifying and weeding out demand easier and more transparent, by making it simple to distinguish between those subjects who are and are not susceptible to demand. The Red Herring technique may also dramatically reduce the number of subjects susceptible to demand, by pulling their suspicions away from the true goals of the study. In our study, 82 subjects bought into our Red Herring on the open-ended Demand Check. These subjects (more than a third of our sample) were interested enough in the study to seek out its “true purpose”. They were also savvy enough to pull out a specific theme other than the rationale they were given. In our study they found the theme of childhood obesity. But if we had not used the Red Herring technique, they might have been more likely to detect the true purpose of the study, which was implanting false memories. If just half of these subjects had found the true goals of the study, we would have doubled the number of subjects susceptible to demand. For example, in another false memory study from our lab, the Red Herring technique was not employed, but subjects were given essentially the same open-ended demand check question at the

end of the study's procedures (Laney, 2006). In this study, a much larger proportion of subjects (20% compared to 8% in the present study) gave "false memory" responses.

There are other questions, however, that the present study could not address. In particular, although some subjects described the evolution of their thought processes about the goals of the study (see example above), most did not. Learning what made subjects suspicious, when they became suspicious, and what they did with those suspicions could help us not only to further identify risks of demand, but also to produce better designed studies in the future.

Study 2

Study 2 looked at the question of demand from a different perspective than that used in Study 1. In particular, it involved a qualitative analysis of subjects' reactions to the false memory study as well as their motivations for responding in the way that they did. A subset of the subjects from Study 1 were interviewed immediately after their participation in that study, but prior to being debriefed. Using a combined approach of qualitative and quantitative methods, we hoped to obtain a fuller picture of subject reactions to experimental stimuli. In particular, in-depth probing from the interviews provided insights on the extent to which demand characteristics were influencing subject responding.

In Study 1, "at risk" subjects were those who indicated that the overall purpose of the study was to create false memories (i.e., those who gave "false memory" responses on the Demand Check). In Study 2 "at risk" subjects were defined in two different ways. Both of these definitions pertained to the study manipulation specifically, rather than the purpose of the study generally. One group of "at risk" subjects (the "Suspicious" group) included those who believed that their profile was fake, whether or not they understood that they were participating in a false memory study. The second group of "at risk" subjects (the "Disagree" group) included those who disagreed with the target asparagus item on their profile.⁴

Method

Subjects

In total, 41 subjects who had completed the Study 1 procedures were recruited and interviewed in Study 2. Only experimental subjects were recruited. A majority of sub-

jects ($n = 31$) were female, and their mean age was 20.3 years (age was missing for seven subjects). As such, these subjects were representative of the full group of experimental subjects in Study 1.

Materials and procedure

All interview subjects completed the main false memory experiment before they were interviewed. All interviews were conducted immediately following data collection from Study 1, but before debriefing. Subjects were interviewed in a different room, close by the experimental lab, and interviews ranged from 15 to 30 min.

Subjects were interviewed individually, and all interviews were semi-structured, with few set questions. Questions addressed subjects' reactions to their profile, their thoughts about the experimental goals and their own behavior. The general question "What did you think of the study?" was asked first to start the interview and put the subject at ease. Subjects were then asked specific questions about their profile, including "What did you think about your profile? How accurate do you think it was? Did it match what you were expecting?" Subjects were also asked about their motivations and thought processes during the experiment [e.g. "Do you think (the profile) affected your responses on the rest of the task?"]. Each subject was fully debriefed at the end of the interview, regarding both the purpose of the interview and (because their original debriefing had been postponed) the original false memory study.

Data coding

Subject responses in the interviews were coded for suspicion and profile disagreement. The traditional concern for demand characteristics in memory distortion research is that subjects may have become suspicious of the cover story, extrapolated the true meaning of the experiment, and pretended to have false memories when in reality they did not. For this reason, responses to the questions regarding the profile items (the Study 1 manipulation) were coded for suspicion. We reasoned that those who failed to recognize the possibility that our manipulation was a manipulation should likewise have failed to intentionally adjust their responses to support the goals of that manipulation.

Subjects were classified as "Suspicious" (as opposed to "Non-suspicious") if they met any of three criteria: (1) they believed that the profiles were probably not "personalized", (i.e., that everyone received the same basic profile), (2) they believed that their profiles had not been generated based on their responses to Session 1 but had been created by the experimenters for some specific purpose, or (3) they thought the profile was intended to influence their

⁴ Disagreement and Suspicion coding were done independently, such that a participant labeled "Disagree" may have been Suspicious or Non-suspicious, and vice versa.

responses in some way during Session 2. Subjects did not have to mention false memories specifically to be classified as Suspicious.

During pilot interviews, the interviewer identified an additional set of subjects who could be susceptible to demand characteristics: those subjects who *disagreed* with their profile feedback. Some subjects, though they believed our story about the source of the profiles, believed one of more of the profile items to be simply inaccurate. That is, they vehemently disagreed with one or more of the items. This disagreement meant that these subjects had to choose between their memories and the content of their profiles when deciding how to answer subsequent relevant questions. Subjects who disagreed with their profile items could either choose to ignore the profile and answer all questions based on their own memories, or they could choose to ignore their real memories and answer as if the profile was correct. The latter option would produce a demand characteristic response, because subjects would be shaping their responses to fit with what they believed the experimenters expected.

A secondary coding scheme was devised in order to classify subjects as either profile “Agrees” or “Disagrees”. Although subjects could have disagreed with any item on their profile, only those who indicated their asparagus feedback item was incorrect were coded as Disagrees for the purpose of this study. When asked why they believed their profile item was wrong, subjects gave a variety of responses ranging from a belief that they had accidentally circled the wrong number on a questionnaire during Session 1 to a belief that the experimenters had given them wrong feedback on purpose. These statements were considered for the Suspicion coding, but not for the Disagree coding; therefore, some subjects who were labeled as “Disagrees” were also labeled as “Suspicious”, but others were not.

Results

Suspicion

During the interview, subjects were either classified as “Suspicious” or “Non-suspicious” about the validity of their feedback profiles. Nine subjects (22%) were classified as Suspicious and 32 (78%) were classified as Non-suspicious.

Mean FHI responses for Suspicious and Non-suspicious subjects can be seen in the upper section of Table 3. Recall that all of the subjects in Study 2 had been *experimental* subjects in Study 1. Suspicious and Non-suspicious subjects had similar mean confidence ratings on the asparagus item at Session 1, and both groups’ confidence increased after the manipulation (Session 2). According to a repeated-

Table 3 Changing confidence (FHI Score) from Study 1 data, split by level of suspicion and whether subjects agreed with their profiles, as determined in Study 2 interviews

	FHI score		Believers
	Session 1	Session 2	
	<i>M (SD)</i>	<i>M (SD)</i>	<i>n (% of group)</i>
<i>Suspicion Level</i>			
Non-suspicious	1.44 (0.88)	2.67 (2.29)	12 (38)
Suspicious	1.75 (1.44)	3.50 (2.38)	2 (22)
<i>Profile Agreement (asparagus item)</i>			
Agree	2.09 (1.56)	4.48 (2.25)	13 (57)
Disagree	1.17 (0.71)	1.83 (1.54)	1 (6)

Suspicious = subjects who (correctly) doubted the motives of the experimenters with respect to the feedback profiles; non-suspicious = subjects who were not suspicious about the source of their profiles; disagree = subjects who believed that the critical item on their profiles was not true of them; agree = subjects who were untroubled by the critical item on their profiles. All subjects were experimental subjects who had completed the Study 1 procedures

measures ANOVA, there was a significant main effect of time, $F(1,39) = 11.66$, $P = 0.002$, but not a significant main effect of suspicion, $F(1,39) = 0.96$, $P = 0.33$, or a significant interaction of suspicion and time, $F(1,39) = 0.37$, $P = 0.55$. This indicates that confidence ratings of suspicious and non-suspicious subjects followed similar patterns, typically starting low, and rising after the manipulation.

We also looked for differences in the proportion of Suspicious and Non-suspicious subjects who met the criteria to be called “believers”. See the right side of Table 3. In results that parallel those of Study 1, we found no significant difference in the proportion of potentially demand characteristic subjects (here Suspicious subjects) and non-demand characteristic subjects who formed false memories $\chi^2(1, n = 41) = 0.39$, $P = 0.33$.

Profile disagreement

Subjects were classified as either profile “Agrees” or “Disagrees”, depending on their reaction to the asparagus item on their profile. Twenty-three subjects (56%) indicated that they agreed with their asparagus feedback, while 18 (44%) were classified as Profile Disagrees.⁵ Subjects who disagreed with their feedback were unequivocal about the discrepancy in their profile: “It said that I totally dislike asparagus, but I only tried asparagus like once in my life and I actually liked it”.

⁵ Of the nine Suspicious participants, just three (33%) were also classified as Disagree participants. Fifteen of the Non-suspicious participants (47%) were classified as Disagrees.

Agree and Disagree subjects' mean FHI responses can be found in the lower section of Table 3. Disagree subjects started out with a lower mean confidence rating in Session 1 than did Agree subjects. This makes sense because it would have helped to drive these subjects' determination that the profile was incorrect. Disagree subjects' ratings also increased less dramatically after the manipulation (Session 2) than did Agree subjects'. This suggests that these subjects, rather than providing demand characteristic responses by trusting the profile over their own memories when judging how to respond to subsequent items, rejected the demand characteristic responses and trusted their own memories when responding. Statistically speaking, a repeated-measures ANOVA gave significant main effects of both time, $F(1,39) = 20.52$, $P < 0.001$, and agreement, $F(1,39) = 20.50$, $P < 0.001$, as well as a significant interaction of time and agreement, $F(1,39) = 6.52$, $P = 0.02$.

Once again, we also looked for differences in the proportion of Agree subjects and Disagree subjects who met the criteria to be called "believers" (see right side of Table 3). In contrast to Suspicious and Non-suspicious subjects, Agree and Disagree subjects behaved differently from each other. While a majority of Agree subjects (57%) met the criteria to be labeled believers, just one Disagree subject (6% of the group) met the criteria. This difference was significant $\chi^2(1, n = 41) = 11.66$, $P = 0.001$.

Subject reports of behavior

In addition to questions about their interpretations of their profiles, subjects were asked whether they believed their profiles had affected their post-manipulation responses. A total of 23 subjects (56%) thought the profile had influenced their later responses on profile-relevant items. The subjects differed, however, in their understanding of this influence. Some subjects (7%) said they used the profile as a guide for recalling what responses they had given for Session 1, while other subjects (7%) reported that the profile made them question their memories for the event items. Nearly a quarter of subjects (22%) reported that they had reacted negatively towards their profiles and thus intentionally overcompensated for the profiles' inaccuracies on their post-manipulation responses by, for example, claiming to be even less confident than they were on the critical items of the post-manipulation FHI. As one subject explained:

I think (the profile affected my answers) because (the feedback was wrong) so I will probably change my answers a little bit to make it a point, you know? ... I gave it a lower number because (my Session 1 responses) made it seem like I like pizza a lot so I thought, ok, I will give it a lower number. I guess I was trying to get my point out.

This behavior clearly demonstrates the defiant type of demand characteristic behavior. That is, the subject concluded that the experimenter wanted him or her to respond in a specific way, and the subject reacted by responding in the opposite way. Additional subjects (12%) reported that the profile affected them in a subtler manner, by making them think more carefully about their responses in an effort to give quality, accurate answers. For example, one subject reported that the profile influenced him by making him try to match his Session 1 and 2 answers exactly, in an effort to make sure the profile items did not bias his answers.

These qualitative reports support the previous statistical analyses that found no evidence of hypothesis-supporting demand characteristic behavior in Study 1. While some subjects reported being persuaded that an FHI item may have happened because they read it on the profile, none reported intentionally giving a high confidence rating to an item simply because they had received feedback about it. If there were any demand characteristics at work, they were of a very different type. Specifically, several subjects stated that they had reacted to incorrect profile items by consciously giving low confidence ratings to those items in order to send a message to the experimenters. This kind of behavior would not help our false memory hypotheses. Instead, it would make our conclusions all the more conservative, because false memories were implanted despite the fact that a sizable proportion of subjects were actively working against our research hypotheses.

Comparing different methods for evaluating demand characteristics

In Study 1, our procedures allowed us to identify a subset of 40 subjects (21% of the sample) who knew (or at least correctly guessed, in the right circumstances) that they were participating in a study of false memory. We then compared the false memory data from these subjects to those from the remaining subjects (who did not guess that they were in a false memory study). We found no significant differences between these two groups. In Study 2, our procedures allowed us to identify a subset of nine subjects (29% of the sample) who were suspicious of their profiles. Again, when these data were compared to those from the remaining, non-suspicious, subjects, no significant differences were found.

The two studies both aimed to identify potentially demand characteristic subjects, but their different methodologies and criteria flagged different groups of subjects for closer inspection. Of the nine subjects who were coded (in Study 2) as being appropriately suspicious of the profile, only three (33%) also chose a "false memory" response in Study 1. While this might lead one to conclude that the interview was a more sensitive test of suspicion, the data show

that seven of the 32 Non-suspicious subjects (22%) from the interview study had in fact marked a “false memory” response earlier during the main false memory experiment.

These results suggest that the different methods were in fact tapping two disparate aspects of suspicion. The questions in the main false memory experiment asked about the perceived general purpose of the experiment, while the interview questions related specifically to perceptions of the profile. Although the two different methodologies employed in Studies 1 and 2 produced two different groups of potentially demand characteristic subjects, neither comparison produced false memory hypothesis-relevant differences between subjects who were and were not “at risk” for demand. That is, any way we looked at it, suspicion had no effect on subject responses.

Discussion

This study used in-depth probing in semi-structured interviews to assess the extent to which the results of a false memory experiment could be explained by demand characteristics. Subjects were categorized along two separate dimensions, suspicion and profile agreement. Suspicious subjects were no more likely to produce false memories than non-suspicious subjects. But those subjects who specifically disagreed with items on their profiles were actually *less* likely to have produced false memories than their fellow subjects.

The results of this study thus indicate that a re-conceptualization of demand characteristic behavior may be useful for false memory studies. In most literature concerned with the issue of experimental demand, the focus has been on those subjects who correctly guess the true purpose of the experiment. However, the Disagree subjects identified in the present study were not particularly suspicious of the study’s procedures [in fact, just three of the 18 Disagree subjects (17%) were also classified as Suspicious], or particularly likely to have guessed the true purpose of the experiment [just four of the 18 Disagree subjects (22%) gave a “false memory” response during the Study 1 procedures]. And yet, Disagree subjects’ experimental results were demonstrably different from those of Agree subjects. Specifically, Disagree subjects’ confidence remained essentially unchanged after the manipulation, and they were extremely unlikely to meet our criteria for having a false memory. This differs markedly from the responses of subjects who agreed with the profile item. Agree subjects’ confidence increased substantially from pre- to post-manipulation and a majority of them met our criteria for having a false memory.

These results suggest that if there is any demand characteristic behavior in our studies, it is likely to run counter to

our hypotheses rather than to favor them. These results also suggest that researchers should expand their focus from merely testing whether subjects were suspicious of their cover story to exploring other possible groups of subjects who may be reacting in unexpected ways to the experimental manipulations. Finally, these results suggest that filler items on study materials (perhaps especially on false feedback profiles) should be carefully tailored to lend credibility to the measures, or there is a risk that these items may actually inoculate subjects against the experimental manipulations.

General discussion

In two studies we attempted to give subjects false memories for two specific food-related childhood events (getting sick after eating asparagus, or loving it the first time they tried it), and then searched for subjects who might have been subject to demand characteristics. We defined susceptibility to demand characteristics in three different ways, and assessed subjects’ behavioral and subjective responses to demand. For each of these definitions, subjects who were classified as “at risk” for demand characteristic responding were in fact no more likely than other subjects to form false memories. For one of the definitions, the “at risk” subjects were actually significantly *less* likely to form false memories. These results suggest that demand characteristics are not a major explanation for our false memory effect.

A major goal of this research was to test a new methodology for minimizing the likelihood of demand in experiments that employ deception. The Red Herring methodology works by giving subjects an outlet for their natural suspicions that is unrelated to the true goals of the experiment. That is, subjects can feel smart for spotting a study theme other than the one they were given, but without actually determining the true nature of the study, and thus becoming subject to demand. When we can be confident that demand is not a major factor in our results, we can be far more confident that those results are meaningful.

The Red Herring methodology can be employed in a variety of psychological experiments where deception is necessary. The Red Herring itself could vary widely, depending on the nature of the study. It need only be plausible (ideally, more plausible than the true purpose of the study), given the materials that subjects encounter during the experimental sessions. Our research has shown that the Red Herring methodology is more successful if it is related directly to the content of materials rather than to their formatting. In a previous (unpublished) study, we modified the formatting of our materials (by printing questionnaires on pink and blue sheets of paper) with minimal content modifications (a few gender-related questions on our standard

questionnaires) to attempt to convince subjects that we were studying gender. This Red Herring was unsuccessful because subjects did not view paper color as related to the purpose of the study. In the present study, our Red Herring (obesity) was related to the content of questionnaires (food and eating), but not to the true purpose of our study (false memory). As such, it worked as intended: It diverted attention away from the true ‘plot’ and convinced even naturally suspicious subjects that the study was about obesity rather than false memory.

Limitations

Although there were traditional experimental and control groups for the false memory manipulation in Study 1, there was no such control group for the Red Herring methodology. That is, all subjects received materials with strategically placed questions about obesity and unhealthy eating. Because these items were fully integrated into the study materials, a Red Herring-free version of the materials would have been substantially different from the standard version.

In addition, the interview methodology employed in Study 2 calls for some caution when interpreting the results. Although care was taken to minimize the time delay between the end of the experiment and the interviews, relying on self-reports has limitations. Research has shown that subjects are often unable to accurately express the true reasons behind their responses (Nisbett & Wilson, 1977). However, demand characteristic behavior has been conceptualized as conscious and willful, and the subject should therefore be cognizant of these motivations when asked about their responses.

Conclusions

Our new Red Herring methodology was successfully employed in Study 1 to divert subjects’ attention from the true goals of the study. It also allowed to us identify those subjects who were particularly likely to respond in a demand characteristic manner, and then analyze their data separately. We used three separate definitions of “at risk” subjects across two different studies with different methodologies. Two of the definitions produced “at risk” subjects who were in fact no more likely to support the hypotheses of our false memory study. The third definition identified a small group of subjects who tended to react to the manipulation by responding in a way that ran counter to the false memory hypotheses. We thus feel confident in concluding that our false memory effects are not due merely to the presence of demand in our studies, and the method described

herein may be useful in the identification of potential demand characteristics in future research.

Acknowledgment This work was partially supported by the Grawemeyer Prize in Psychology, awarded to Elizabeth Loftus. In addition, we would like to thank the other members of the “Memory & Law” seminar at the University of California, Irvine where the Red Herring idea first took shape.

References

- Bernstein, D. M., Laney, C., Morris, E. K., & Loftus, E. F. (2005a). False memories about food can lead to food avoidance. *Social Cognition*, *23*, 10–33.
- Bernstein, D. M., Laney, C., Morris, E. K., & Loftus, E. F. (2005b). False beliefs about fattening foods can have healthy consequences. *Proceedings of the National Academy of Sciences*, *102*, 13724–13731.
- Bjorklund, D. F., Cassel, W. S., Bjorklund, B. R., Brown, R. D., Park, C. L., et al., (2000). Social demand characteristics in children’s and adults’ eyewitness memory and suggestibility: the effects of different interviews on free recall and recognition. *Applied Cognitive Psychology*, *14*, 421–433.
- Bryant, R. A., & Mallard, D. (2005). Reality monitoring in hypnosis: a real-simulating analysis. *International Journal of Clinical and Experimental Hypnosis*, *53*, 13–25.
- Bryant, R. A., Barnier, A. J., Mallard, D., & Tibbits, R. (1999). Post-hypnotic amnesia for material learned before hypnosis. *International Journal of Clinical and Experimental Hypnosis*, *47*, 46–64.
- Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review*, *3*, 208–214.
- Hyman, I. E. Jr, Husband, T. H., & Billings, F. J. (1995). False memories of childhood experiences. *Applied Cognitive Psychology*, *9*, 181–197.
- Lampinen, J. M., Neuschatz, J. M., & Payne, D. G. (1999). Source attributions and false memories: a test of the demand characteristics account. *Psychonomic Bulletin & Review*, *6*, 130–135.
- Laney, C. (2006). *Emotional content of true and false memories*. Unpublished doctoral dissertation, University of California, Irvine.
- Lindsay, D. S. (1990). Misleading suggestions can impair eyewitnesses’ ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 1077–1083.
- Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, *25*, 720–725.
- Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 19–31.
- MacLeod, C. M. (1999). The item and list methods of directed forgetting: test differences and the role of demand characteristics. *Psychonomic Bulletin & Review*, *6*, 123–129.
- Morris, E. K., Laney, C., Bernstein, D. M., & Loftus, E. F. (2006). Susceptibility to memory distortion: how do we decide it has occurred? *American Journal of Psychology*, *119*, 255–276.
- Nisbett, R., & Wilson, T. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review*, *84*, 231–259.
- Orne, M. T. (1959). The nature of hypnosis: artifact and essence. *Journal of Abnormal and Social Psychology*, *58*, 277–299.
- Orne, M. T. (1962). On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications. *American Psychologist*, *17*, 776–783.

- Orne, M. T. (1979). On the simulating subjects as a quasi-control group in hypnosis research: What, why, and how. In: E. Fromm, & R. E. Shor (Eds.), *Hypnosis: Developments in Research and New Perspectives*, 2nd edn (pp. 519–565). Chicago: Aldine.
- Page, M. M., & Scheidt, R. J. (1971). The elusive weapons effect: demand awareness, valuation apprehension, and slightly sophisticated subjects. *Journal of Personality and Social Psychology*, *20*, 304–318.
- Wade, K. A., Garry, M., Read, J. D., & Lindsay, D. S. (2002). A picture is worth a thousand lies: using false photographs to create false childhood memories. *Psychonomic Bulletin & Review*, *9*, 597–603.
- Weinberg, H. I., Wadsworth, J., & Baron, R. S. (1983). Demand and the impact of leading questions on eyewitness testimony. *Memory & Cognition*, *11*, 101–104.