

SAGE BENCHMARKS IN PSYCHOLOGY

**THEORETICAL
PSYCHOLOGY –
CONTEMPORARY
READINGS**

VOLUME II

Theory and Method

Edited by

Henderikus J. Stam

 **SAGE**

Los Angeles | London | New Delhi
Singapore | Washington DC

58/2012 / 20813-2

© Introduction and editorial arrangement by Henderikus J. Stam, 2012

First published 2012

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act, 1988, this publication may be reproduced, stored or transmitted in any form, or by any means, only with the prior permission in writing of the publishers, or in the case of reprographic reproduction, in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

Every effort has been made to trace and acknowledge all the copyright owners of the material reprinted herein. However, if any copyright owners have not been located and contacted at the time of publication, the publishers will be pleased to make the necessary arrangements at the first opportunity.

SAGE Publications Ltd
1 Oliver's Yard
55 City Road
London EC1Y 1SP

SAGE Publications Inc.
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
B 1/1 I, Mohan Cooperative Industrial Area
Mathura Road
New Delhi 110 044

SAGE Publications Asia-Pacific Pte Ltd
3 Church Street
#10-04 Samsung Hub
Singapore 049483

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

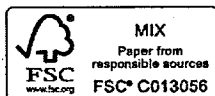
ISBN: 978-1-84920-773-7 (set of four volumes)

Library of Congress Control Number: 2011925392

Typeset by Star Compugraphics Private Limited, Delhi

Printed on paper from sustainable resources

Printed and bound in Great Britain by TJ International Ltd, Padstow, Cornwall



Experimental Practices in Economics: A Methodological Challenge for Psychologists?

Ralph Hertwig and Andreas Ortmann

1. Introduction

Empirical tests of theories depend crucially on the methodological decisions researchers make in designing and implementing the test (Duhem 1953; Quine 1953). Analyzing and changing specific methodological practices, however, can be a challenge. In psychology, for instance, “it is remarkable that despite two decades of counterrevolutionary attacks, the mystifying doctrine of null hypothesis testing is still today the Bible from which our future research generation is taught” (Gigerenzer & Murray 1987, p. 27). Why is it so difficult to change scientists’ practices? One answer is that our methodological habits, rituals, and perhaps even quasi-religious attitudes about good experimentation are deeply entrenched in our daily routines as scientists, and hence often not reflected upon.

To put our practices into perspective and reflect on the costs and benefits associated with them, it is useful to look at methodological practices across time or across disciplines. Adopting mostly the latter perspective, in this article we point out that two related disciplines, experimental economics and corresponding areas in psychology (in particular, behavioral decision making) have very different conceptions of good experimentation.

We discuss the different conceptions of good experimentation in terms of four key variables of experimental design and show how these variables tend to be realized differently in the two disciplines. In addition, we show that experimental standards in economics, such as performance-based monetary payments (henceforth, *financial incentives*) and the proscription against deception, are rigorously enforced through conventions or third parties. As a result, these standards allow for little variation in the experimental practices of individual researchers. The experimental standards in psychology, by contrast, are comparatively *laissez-faire*, allowing for a wider range of practices. The lack of procedural regularity and the imprecisely specified social situation "experiment" that results may help to explain why in the "muddy vineyards" (Rosenthal 1990, p. 775) of soft psychology, empirical results "seem ephemeral and unreplicable" (p. 775).

1.1. The Uncertain Meaning of the Social Situation "Experiment"

In his book on the historical origins of psychological experimentation, Danziger (1990) concluded that "until relatively recently the total blindness of psychological investigators to the social features of their investigative situations constituted one of the most characteristic features of their research practice" (p. 8). This is deplorable because the experimenter and the human data source are necessarily engaged in a social relationship; therefore, experimental results in psychology will always be codetermined by the social relationship between experimenter and participant. Schultz (1969) observed that this relationship "has some of the characteristics of a superior-subordinate one. . . . Perhaps the only other such one-sided relationships are those of parent and child, physician and patient, or drill sergeant and trainee" (p. 221). The asymmetry of this relationship is compounded by the fact that the experimenter knows the practices of experimentation by virtue of training and experience, while the typical subject is participating in any given experiment for the first time.¹

Under these circumstances, and without clear-cut instructions from the experimenter, participants may generate a variety of interpretations of the experimental situation and therefore react in diverse ways to the experimental stimuli. In the words of Dawes (1996):

The objects of study in our experiments (i.e., people) have desires, goals, pre-suppositions, and beliefs about what it is we wish to find out. Only when it is explicitly clear that what we are seeking is maximal performance . . . can we even safely assume that our interpretation of the experimental situation corresponds to that of our subjects. . . . Even then, however, we may not be able to . . . "control for" factors that are not those we are investigating. (p. 20)

1.2. Defining the Social Situation "Experiment"

In this article, we argue that experimental standards in economics reduce participants' uncertainty because they require experimenters to specify precisely the "game or contest" (Rieken 1962, p. 31) between experimenter and participant in a number of ways. In what follows, we consider four key features of experimental practices in economics, namely, script enactment, repeated trials, financial incentives, and the proscription against deception. The differences between psychology and economics on these four features can be summed up – albeit in a simplified way – as follows. Whereas economists bring a precisely defined "script" to experiments and have participants enact it, psychologists often do not provide such a script. Economists often repeat experimental trials; psychologists typically do not. Economists almost always pay participants according to clearly defined performance criteria; psychologists usually pay a flat fee or grant a fixed amount of course credit. Economists do not deceive participants; psychologists, particularly in social psychology, often do.

We argue that economists' realizations of these variables of experimental design reduce participants' uncertainty by explicitly stating action choices (script), allowing participants to gain experience with the situation (repeated trials), making clear that the goal is to perform as well as they can (financial incentives), and limiting second-guessing about the purpose of the experiment (no deception). In contrast, psychologists' realizations of these variables tend to allow more room for uncertainty by leaving it unclear what the action choices are (no script), affording little opportunity for learning (no repeated trials), leaving it unclear what the experimenters want (no financial incentives), and prompting participants to second-guess (deception).

Before we explore these differences in detail, four caveats are in order. First, the four variables of experimental design we discuss are, in our view, particularly important design variables. This does not mean that we consider others to be irrelevant. For example, we question economists' usual assumption that the abstract laboratory environment in their experiments is neutral and, drawing on results from cognitive psychology, have argued this point elsewhere (Ortmann & Gigerenzer 1997). Second, we stress that whenever we speak of standard experimental practices in "psychology," we mean those used in research on behavioral decision making (an area relevant to both psychologists and economists; e.g., Rabin 1998) and related research areas in social and cognitive psychology such as social cognition, problem solving, and reasoning. The practices discussed and the criticisms leveled here do not apply (or do so to a lesser degree), for instance, to research practices in sensation and perception, biological psychology, psycho-physics, learning, and related fields. Third, we do not provide an exhaustive review of the relevant literature, which, given the wide scope of the paper, would have been

a life's work. Rather, we use examples and analyze several random samples of studies to show how differences in the way the design variables are realized can affect the results obtained. Moreover, even in discussing the limited areas of research considered here, we are contrasting prototypes of experimental practices to which we are aware many exceptions exist.

Finally, we do not believe that the conventions and practices of experimental economists constitute the gold standard of experimentation. For example, we concur with some authors' claim that economists' strict convention of providing financial incentives may be too rigid and may merit reevaluation (e.g., Camerer & Hogarth 1999). The case for such reevaluation has also been made in a recent symposium in *The Economic Journal* (e.g., Loewenstein 1999). This symposium also takes issue with the assumed neutrality of the laboratory environment (e.g., Loomes 1999), scripts that are too detailed (e.g., Binmore 1999; Loewenstein 1999; Loomes 1999; Starmer 1999), and the relevance of one-shot decision making (e.g., Binmore 1999; Loewenstein 1999), among other aspects of experimentation in economics that warrant reevaluation (e.g., Ortmann & Tichy 1999). In other words, a paper entitled "Experimental practices in psychology: A challenge for economists?" may well be worth writing.

2. Enacting a Script versus "Ad-libbing"

Economists run experiments usually for one of three reasons: to test decision-theoretic or game-theoretic models, to explore the impact of institutional details and procedures, or to improve understanding of policy problems such as the behavior of different pricing institutions (e.g., Davis & Holt 1993, Ch. 3 on auctions and Ch. 4 on posted offers).

To further understanding of policy problems, experimental economists construct small-scale abstractions of real-world problems (although typically these miniature replicas are framed in abstract terms). To test theoretical models, economists attempt to translate the model under consideration into a laboratory set-up that is meant to capture the essence of the relevant theory. This mapping inevitably requires the experimenter to make decisions about "institutional details" (i.e., the degree of information provided in the instructions, the way the information is presented to participants, the communication allowed between participants, etc.). Economists have learned to appreciate the importance of such institutional details and procedures, and how these might affect results (e.g., Davis & Holt 1993, pp. 507–509; Osborne & Rubinstein 1990; Zwick et al. 1999).

To enhance replicability and to trace the sometimes subtle influence of institutional details and experimental parameters, experimental economists have come to provide participants with scripts (instructions) that supply descriptions of players, their action choices, and the possible payoffs (for standard examples of such instructions, see appendices in Davis & Holt 1993).

Economists then ask participants to enact those scripts. For example, they assign each of them the role of buyer or seller and ask them to make decisions (e.g., to buy or sell assets) that determine the amount they are paid for their participation, a practice discussed in detail later.

An example of a script and its enactment is provided by Camerer et al. (1989) in their investigation of hindsight bias.² In their design, an "uninformed" group of participants guessed future earnings of real companies based on information such as the previous annual earnings per share. An "informed" group of participants (who were told the actual earnings) then traded assets that paid dividends equal to the earnings predicted by the uninformed group. Participants in both groups were provided with a precise script. Those in the uninformed group were given the role (script) of a market analyst faced with the task of predicting the future dividends of various companies. Those in the informed group were assigned the role of trader: they knew that the dividend was determined by the uninformed group's predictions. Thus, to price the assets optimally (and thereby to avoid hindsight bias), the "traders" had to predict the prediction of the "analysts" accurately, that is, to ignore their knowledge of the actual dividends. Eventually, the traders traded the assets to others in actual double-oral auctions, in which "buyers and sellers shouted out bids or offers at which they were willing to buy or sell. When a bid and offer matched, a trade took place" (p. 1236).

Unlike Camerer et al.'s (1989) study, typical hindsight bias experiments in psychology do not provide participants with a script, thus forcing them to ad-lib, that is, to infer the meaning of the experiment as they go. In a typical study (Davies 1992), participants were given a series of assertions and asked to rate the truth of each. They were then given feedback (i.e., the truth values of the assertions) and later asked to recall their original judgment. In contrast to Camerer et al. (1989), Davies did not assign specific roles to participants or provide them with any precise script. Instead, the first stage of the study, during which participants rated assertions for their truth, was merely described to participants as "involving evaluation of college students' knowledge" (Davies 1992, p. 61), and they were told that the recollection stage "concerned people's ability to remember or recreate a previous state of knowledge" (Davies 1992, p. 61). This procedure is typical of many psychological studies on hindsight bias (e.g., Hell et al. 1988; Hoffrage & Hertwig 1999).

In psychological research on judgment, decision making, and reasoning, too, researchers typically do not provide participants with a script to enact. Much of this research involves word problems such as the conjunction task (e.g., Tversky & Kahneman 1983), the engineer-lawyer task (e.g., Kahneman & Tversky 1973), the Wason selection task (e.g., Evans et al. 1993), and the 2-4-6 task (e.g., Butera et al. 1996). These problems share a number of typical features. For example, they often are ambiguous (e.g., use polysemous terms such as "probability," see Hertwig & Gigerenzer 1999) and require participants to ignore conversational maxims in order to reach the "correct" solution (see Hilton 1995).³ Furthermore, they do not require participants to assume

clearly specified roles, like the analysts and traders in Camerer et al.'s (1989) study, or to enact a script. As a result, participants are forced to ad-lib.

Participants' ad-libbing is likely to be influenced by their expectations about what experimenters are looking for. Providing a script would not alter the fact that the typical participant in psychology (and economics) has never or rarely encountered a particular experimental situation before. That is, notwithstanding provision of a script, participants are still likely to be sensitive to cues that are communicated to them by means of campus scuttlebutt, the experimenter's behavior, and the research setting. However, scripts can constrain participants' interpretations of the situation by focusing their attention on those cues that are intentionally communicated by the experimenter (e.g., the task instructions), thus clarifying the demand characteristics of the social situation "experiment." As a consequence, scripts may enhance replicability.

Enacting a script is closely related to "role playing" in social psychology (e.g., Greenwood 1983; Krupat 1977), in which the "intent is for the subject to directly and actively involve himself in the experiment, and to conscientiously participate in the experimental task" (Schultz 1969, p. 226). To borrow the terminology of Hamilton's useful three-dimensional classification (referred to in Geller 1978, p. 221), the role-playing simulations that come closest to economics experiments are those performed (rather than imagined) and scripted (rather than improvised), and in which the dependent variable is behavior (rather than verbal utterances). In economics experiments, however, participants do not just simulate but are real agents whose choices have tangible consequences for them. For example, in the Camerer et al. (1989) study, they were real analysts and real traders, albeit in a scaled-down version of a real market.

2.1. Does Providing and Enacting a Script Matter?

We believe that providing a script for participants to enact affects experimental results. At the same time, we readily admit that the evidence for this claim is at present tenuous because provision of scripts and their enactment are rarely treated as independent variables. Using as examples the prediction task in Camerer et al.'s (1989) study and the Wason selection task in psychology, we now discuss the potential importance of providing a script and having participants enact it.

Camerer et al. (1989) compared the amount of hindsight bias in the predictions of participants who enacted the role of trader (i.e., who actually traded assets in the double-oral auction) to the bias in predictions made by another group of participants who did not enact the role of trader. The goal of the two groups was the same: to predict the average prediction of the uninformed group, given companies' actual earnings. Both groups received incentives for making correct predictions. Camerer et al. (1989) reported that participants in both conditions exhibited some hindsight bias, but enactment of the trader

role reduced the bias by about half: The difference in hindsight bias between the two groups was $r = .18$ (calculated from data in their Fig. 4), a small to medium effect (see Rosenthal & Rosnow 1991, p. 444).

Research on the Wason selection task provides another example of a situation in which providing a script (or more precisely, a proxy for one) – namely, assigning participants to the perspective of a particular character – dramatically changes their responses. This task is perhaps the most studied word problem in cognitive psychology. In what is known as its abstract form, participants are shown four cards displaying symbols such as *T*, *J*, *4*, and *8* and are given a conditional rule about the cards, such as “If there is a *T* on one side of the card [antecedent *P*], then there is a *4* on the other side of the card [consequent *Q*].” Participants are told that each card has a letter on one side and a number on the other. They are then asked which cards they would need to turn over in order to discover whether the conditional rule is true or false. The typical result, which has been replicated many times (for a review, see Evans et al. 1993, Ch. 4), is that very few participants (typically only about 10%) give the answer prescribed by propositional logic: *T* and *8* (*P* & *not-Q*). Most participants choose either *T* (*P*) alone or *T* and *4* (*P* & *Q*). These “errors” in logical reasoning have been seen as reflections of the confirmation bias, the matching bias, and the availability heuristic (for a review, see Garnham & Oakhill 1994).

The original, abstract Wason selection task was content-free. Numerous researchers have since shown that dressing it in thematic garb, that is, putting it in a social context, increases the percentage of logically correct answers. In one such task, a police officer is checking whether people conform to certain rules: in the context of a drinking age law (“If someone is drinking beer [*P*], then they must be over 19 years of age [*Q*]”), 74% of participants gave the logical *P* & *not-Q* response (Griggs & Cox 1982). Gigerenzer and Hug (1992) later demonstrated that the way in which social context affects reasoning in the selection task also depends on the *perspective* into which participants are cued. For instance, the implications of the rule “If an employee works on the weekend, then that person gets a day off during the week” depend on whether it is seen from the perspective of an employer or of an employee. Among participants cued into the role of an employee, the dominant answer was *P* & *not-Q* (75%); among participants cued into the role of an employer, in contrast, the dominant response was *not-P* & *Q* (61%; for more detail, see Ortmann & Gigerenzer 1997). Perspective can thus induce people to assume certain social roles, activating a script like those provided in economics experiments.⁴

To conclude this section, the effects of role playing in Camerer et al.’s (1989) study and perspective taking in selection tasks suggest that supplying a script for participants to enact can make an important difference to the results obtained. Although script provision (i.e., action choices, payoffs, perspective, etc.) demands more elaborate and transparent instructions (e.g., compare Camerer et al.’s market study with any typical hindsight bias study

in psychology), it is likely to reduce the ambiguity of the experimental situation and thereby increase researchers' control over participants' possible interpretations of it. This practice is also likely to enhance the replicability of experimental results. We propose that psychologists consider having participants enact scripts wherever possible.

3. Repeated Trials versus Snapshot Studies

Economists use repeated trials for (at least) two reasons. The first is to give participants a chance to adapt to the environment, that is, to accrue experience with the experimental setting and procedure. This motivation applies to both decision and game situations and reflects economists' interest in the impact of experience on behavior. Binmore (1994) articulated this rationale as follows:

But how much attention should we pay to experiments that tell us how inexperienced people behave when placed in situations with which they are unfamiliar, and in which the incentives for thinking things through carefully are negligible or absent altogether? . . . Does it [the participant's behavior] survive after the subjects have had a long time to familiarize themselves with all the wrinkles of the unusual situation in which the experimenter has placed them? If not, then the experimenter has probably done no more than inadvertently trigger a response in the subjects that is adapted to some real-life situation, but which bears only a superficial resemblance to the problem the subjects are really facing in the laboratory. (pp. 184–85)

The second motivation for the use of repeated trials, while also reflecting economists' interests in the impact of experience on behavior, is specific to game situations. Repeated trials afford participants the opportunity to learn how their own choices interact with those of other players in that specific situation. Although in practice the two kinds of learning are difficult to distinguish, they are conceptually distinct. The first kind of learning (adapting to the laboratory environment) relates to a methodological concern that participants may not initially understand the laboratory environment and task, whereas the second kind of learning (understanding how one's own choices interact with those of other participants) relates to the understanding of the possibly strategic aspects of the decision situation. Game theory captures those strategic aspects and suggests that for certain classes of games, people's behavior "today" will depend on whether and how often they may be paired with others in the future.

Underlying both motivations for the use of repeated trials is economists' theoretical interest in equilibrium solutions, that is, the hope that for every scenario a belief or behavior exists that participants have no incentive to change. However, equilibrium is assumed not to be reached right away. Rather, it is expected to evolve until participants believe their behavior to be optimal for the situation they have been placed in. This is why in economics

experiments “special attention is paid to the last periods of the experiment . . . or to the change in behavior across trials. Rarely is rejection of a theory using first-round data given much significance” (Camerer 1997, p. 319). Note, however, that although economists tend to use repeated trials most of the time, there are important exceptions. For instance, most studies of trust games (e.g., Berg et al. 1995), dictator games (e.g., Hoffman et al. 1996), and ultimatum games employ one-shot scenarios. It is interesting to consider whether the attention-grabbing results of these games are due to the very fact that they are typically implemented as one-shot rather than repeated games.

Typically, economists implement repeated trials either as *stationary replications* of one-shot decision and game situations or as repeated game situations. Stationary replication of simple decision situations (i.e., without other participants) involves having participants make decisions repeatedly in the same one-shot situation. Stationary replication of game situations also involves having participants make decisions repeatedly in the same one-shot situation, but with new participants in each round. In contrast, other repeated game situations may match participants repeatedly with one another and thus allow for strategic behavior. Neither stationary replication of one-shot decision and game situations nor other repeated game situations implement environments that change. Instead, learning is typically studied in environments whose parameterization (e.g., payoff structure) does not change. Camerer (1997) referred to such situations as “Groundhog Day replication” (p. 319). In what follows, we focus on the special case of Groundhog Day replication referred to as stationary replication above.

In contrast to economists, researchers in behavioral decision making typically provide little or “no opportunity for learning” (Thaler 1987, p. 119; see also Hogarth 1981; Winkler & Murphy 1973), tending instead to conduct “snapshot” studies. It would be misleading, however, to suggest that psychologists have ignored the role of feedback and learning. For instance, there is a history of multi-stage decision making in research on behavioral decision making (see Rapoport & Wallsten 1972). Moreover, studies in which repetition and feedback are used can be found in research on multiple-cue probability learning (e.g., Balzer et al. 1989; Klayman 1988), social judgment theory (Hammond et al. 1975), dynamic decision making (e.g., Brehmer 1992; 1996; Diehl & Serman 1995; Edwards 1962), probabilistic information processing (e.g., Wallsten 1976), and in research on the effects of different kinds of feedback (e.g., Creyer et al. 1990; Hogarth et al. 1991). Nevertheless, “most judgment research has focused on discrete events. This has led to underestimating the importance of feedback in ongoing processes” (Hogarth 1981, p. 197).

To quantify the use of repeated trials and feedback in behavioral decision making, we analyzed a classic area of research in this field, namely, that on the base-rate fallacy. For the last 30 years, much research has been devoted to the observation of “fallacies,” “biases,” or “cognitive illusions” in inductive

reasoning (e.g., systematic deviations from the laws of probability). Among them, the base-rate fallacy "had a celebrity status in the literature" (Koehler 1996, p. 2). According to Koehler's (1996) recent review of base-rate fallacy research, "hundreds of laboratory studies have been conducted on the use of base rates in probability judgment tasks" (p. 2), and "investigators frequently conclude that base rates are universally ignored" (p. 2). How many of these laboratory studies have paid attention to the possible effects of feedback and learning?

To answer this question, we examined the articles cited in Koehler's (1996) comprehensive review of Bayesian reasoning research. We included in our analysis all empirical studies on the use of base rates published in psychology journals (excluding journals from other disciplines and publications other than articles) since 1973, the year in which Kahneman and Tversky published their classic study on the base-rate fallacy. This sample comprises a variety of paradigms, including, for instance, word problems (e.g., engineer-lawyer and cab problems), variations thereof, and "social judgment" studies (which explore the use of base rates in social cognition such as stereotype-related trait judgments). As the unit of analysis, we took studies – most articles report more than one – in which an original empirical investigation was reported.

By these criteria, 106 studies were included in the analysis. Although this sample is not comprehensive, we believe it is representative of the population of studies on the use of base rates. Of the 106 studies, only 11 (10%) provided participants with some kind of trial-by-trial feedback on their performance (Study 1 in Manis et al. 1980; Studies 1 and 2 in Lopes 1987; Lindeman et al. 1988; Studies 1–5 in Medin & Edelson 1988; Studies 1 and 2 in Medin & Bettger 1991). The picture becomes even more extreme if one considers only those studies that used (sometimes among others) the classic word problems (engineer-lawyer and cab problem) employed by Kahneman and Tversky (1973) or variants thereof. Among these 36 studies, only 1 provided trial-by-trial feedback concerning participants' posterior probability estimates (Lindeman et al. 1988). Based on this survey, we conclude that repetition and trial-by-trial feedback is the exception in research on the base-rate fallacy. This conclusion is consistent with that drawn by Hogarth (1981) almost 20 years ago, namely, that "many discrete judgment tasks studied in the literature take place in environments degraded by the lack of feedback and redundancy. . . . As examples, consider studies of Bayesian probability revision" (p. 199).

3.1. Do Repetition and Feedback Matter?

There is evidence from economists' research on the use of base rates involving repeated trials that they do, indeed. When trials are repeated, base rates do not seem to be universally ignored. Harrison (1994) designed an experiment

to test, among other things, the effect of repetition (plus feedback) and the validity of the representativeness heuristic, which Kahneman and Tversky (1973) proposed as an explanation for people's "neglect" of base rates. This explanation essentially states that people will judge the probability of a sample by assessing "the degree of correspondence [or similarity] between a sample and a population" (Tversky & Kahneman 1983, p. 295).

Unlike Kahneman and Tversky (1973), but like Grether (1980; 1992), Harrison used a bookbag-and-poker-chips paradigm in which participants had to decide from which of two urns, A and B, a sample of six balls (marked with either Ns or Gs) had been drawn. In addition to the ratio of Ns and Gs in the sample and the frequencies of Ns and Gs in the urns (urn A: four Ns and two Gs, urn B: three Ns and three Gs), participants knew the urns' priors (i.e., the probabilities with which each of the two urns was selected). In this design, the ratio of Ns and Gs in the sample can be chosen so that use of the representativeness heuristic leads to the choice of urn A (as the origin of the sample of six balls), whereas application of Bayes's theorem leads to the choice of urn B, and vice versa.

Participants in Harrison's (1994) study judged a total of 20 samples. After each one, participants were told from which urn the balls were drawn. After each set of 10 decisions, their earnings were tallied based on the number of choices they made in accordance with Bayes's theorem. There were three payoff schedules: Two were contingent on performance and one was not. Harrison (1994) split the choices according to whether they were made when participants were "inexperienced" (first set of 10 decisions) or "experienced" (second set of 10 decisions). He found that the representativeness heuristic strongly influenced the decisions of participants who were inexperienced and unmotivated, that is, who had completed only the first set of 10 decisions and who received a fixed amount of money (independent of performance). However, he also found that when those participants who were not monetarily motivated made the *second* set of 10 decisions, "the Heuristic has no noticeable influence at all" (pp. 249–50). Moreover, Harrison (1994) reported finding little to no evidence of the representativeness heuristic among inexperienced participants (i.e., in the first set of 10 decisions) whose earnings were based on performance.

Harrison's (1994) results seem to contradict Grether's (1980). Grether concluded that participants do tend to follow the representativeness heuristic. However, Grether employed a different definition of experience. Specifically, he counted every participant who had previously assessed the *same* prior-sample combination as experienced. In Harrison's study, in contrast, participants had to make 10 judgments with feedback before they were considered experienced. That experience can substantially improve Bayesian reasoning has also been shown in a series of experiments by Camerer (1990); he also observed that the significance of the biases increased because the variance decreased with experience. The three studies taken together strongly suggest that one

ought to use repeated trials when studying Bayesian reasoning, and that biases diminish in magnitude with sufficient experience (Camerer 1990; Harrison 1994), although not necessarily after only a few trials (Grether 1980).

This conclusion is also confirmed by a set of experiments conducted in psychology. In Wallsten's (1976) experiments on Bayesian revision of opinion, participants completed a large number of trials. In each trial, participants observed events (samples of numbers), decided which of two binomial distributions was the source, and estimated their confidence in the decision. Participants received trial-by-trial feedback, and the sampling probabilities of the two populations under consideration changed from trial to trial. The results showed strong effects of experience on Bayesian reasoning. In the early trials, participants tended to ignore the sampling probability under the less likely hypothesis. As they gained experience, however, they increasingly gave more equal weight to the likelihood of the data under each of the two hypotheses (also see Wallsten 1972).

What are the results in the few studies in our sample that examined the use of base rates using trial-by-trial feedback? Only 4 of these 11 studies (Lindeman et al. 1988; Lopes 1987; Manis et al. 1980) systematically explored the effect of repetition and feedback by comparing a feedback and a no-feedback condition. Table 1 summarizes the results of these four studies. Although the small sample size limits the generalizability of the findings, the results in Table 1 indicate that providing people with an opportunity to learn does increase the extent to which base rates are used, and does bring Bayesian inferences closer to the norm.

However, cautionary notes are in order: Manis et al.'s findings have been suggested to be consistent with reliance on representativeness (Bar-Hillel & Fischhoff 1981); in Lindeman et al.'s (1988) study, the effect of learning did not generalize to a new problem (which according to Lindeman et al. could be due to a floor effect), and in Lopes's (1987) studies the effects of performance-dependent feedback and a training procedure cannot be separated. More generally, Medin and Edelson (1988) caution that people's use of base-rate information "must be qualified in terms of particular learning strategies, category structures, and types of tests" (p. 81).

In the same sample of studies, we also found some that investigated the effect on Bayesian reasoning of "mere practice," that is, the use of repeated trials without feedback. According to these studies, even mere practice can make a difference. With repeated exposure, it seems that "respondents tended to be influenced by the base rate information to a greater degree" (Hinsz et al. 1988, p. 135; see also Fischhoff et al. 1979, p. 347). Moreover, mere practice seems to increase slightly the proportion of Bayesian responses (Gigerenzer & Hoffrage 1995), and can increase markedly participants' consistency (i.e., in applying the same cognitive algorithm across tasks). Mere practice also may drastically alter the distribution of responses: "in the one-judgment task, subjects appear to respond with one of the values given, whereas when

Table 1: Effects of trial-by-trial feedback on the use of base rates obtained in studies from Koehler's (1996) review

	Task/Response/Feedback	Comparison	General results
Manis, Dovalina, Avis & Cardoze (1980)	<p>Task. In each of 50 trials, participants viewed a photograph of a young male, guessed his attitude on an issue such as legalization of the use of marijuana.</p> <p>Response. Binary classification (pro vs. con).</p> <p>Feedback. After each trial, feedback on binary classification.</p>	<p>Study 1 Two feedback conditions. 80% vs. 20% "pro" attitude base rates Control condition. No feedback</p>	<p>"Base-rate information had a clear-cut effect on predictive behavior" (p. 235). By the end of the 50 trials, respondents in the no-feedback condition predicted that about half of the pictured men have a "pro" attitude, whereas in the two feedback conditions, the average percentage of predicted "pro" attitude was 70% and 30% in the 80% and 20% base-rate conditions, respectively.</p>
Lopes (1987)	<p>Task. In each of 175 trials, participants were required to decide from which of two alternative Bernoulli processes a sample was generated.*</p> <p>Response. Rating as to whether the sample comes from process 1 or 2.</p> <p>Feedback. On 13 practice and training trials, participants received training and feedback on a judgment operation.</p>	<p>Study 1 Training condition Control condition Received no training</p>	<p>Trained participants made fewer "adjustment errors" than untrained participants for "strong-weak" pairs ($\epsilon_{tw} = .84$) and diagonal pairs ($\epsilon_{tt} = .42$) but not for "weak-strong" pairs, and were more accurate in their judgments: Root-mean-squared deviations between obtained and optimal judgments were .07 and .02 for untrained participants, and .02 and .005 for trained participants, respectively. However, training did not make participants' responses qualitatively more Bayesian (p. 174).</p>
	<p>Same task but modified training (182 trials, of which 20 were practice and feedback trials)</p>	<p>Study 2 Training condition Control condition</p>	<p>"Clearly, the training procedure has been effective in bringing subjects' responses closer to optimal" (p. 178). Root-mean-squared deviations between obtained and optimal response were .05 and .10 for trained and untrained participants, respectively. The optimal response requires an interaction effect that accounts for 4.66% of the systematic sums of square (in terms of analysis of variance). In the training condition, this value is 4.55%, whereas in the control condition the value is .42%.</p>

(Continued)

Table 1: (Continued)

	Task/Response/Feedback	Comparison	General results
Lindeman, van den Brink & Hoogstraten (1988)	<p>Task. In each of 16 trials, participants predicted the probability that a person described in a personality sketch has a certain profession (engineer-lawyer problem: two different base rates per description); after the training phase all participants were presented with a new problem (divorce problem)</p> <p>Response. Probability estimate</p> <p>Feedback. After the second estimate per description, feedback on the required probability estimate for second estimate</p>	<p>Feedback</p> <p>Training-only</p> <p>32 probability estimates but no feedback</p> <p>No treatment</p> <p>No training</p>	<p>For the engineer-lawyer problem (the training problem), "over-all performance of subjects who received feedback was closer to the Bayesian norm than the performance of subjects who received training only" (p. 346). In the training-only condition, 50% of the mean differences between obtained and Bayesian probability estimates were significant (at $p = .05$); in the feedback condition, in contrast, only 13% of the mean differences were significant.</p> <p>Although the positive effects of feedback "do not generalize to" (p. 349) the divorce problem, thinking-aloud protocols show that base rates were less often mentioned in the no-treatment condition ($n = 3$ out of 16 protocols, 19%) than in the feedback condition ($n = 11$ out of 18 protocols, 61%)^b.</p>

^aThe context story asked participants to imagine making judgments concerning the maintenance of milling machines. The judgment concerns whether or not a crucial part has broken inside the machine. For each machine, participants received two samples; there were three kinds of pairs of samples (weak-strong, strong-weak, and diagonal); only the latter two are test cases for the training as here intuitive and normative response seem to diverge). In Study 1, the training procedure taught participants only that adjustment of the initial rating made after presentation of the second sample should always be in the direction of the hypothesis favored by the second sample. In Study 2, the training also taught participants to process the two samples in the order of their apparent relative strength.

^bAfter the training phase, participants in all three conditions were presented with a different problem. Lindeman et al. (1988) speculated that the failure to find an effect of training for the estimates on this problem may be related to a floor effect: "the divorce problem is inherently easier than the training problem, so that subjects might get the test item right even without the training phase" (p. 349).

given many problems, they appear to integrate the information" (Birnbbaum & Mellers, 1983 p. 796).

Taken together, these examples illustrate that repetition of trials combined with performance feedback, and to some extent even mere practice (repetition without feedback), can improve participants' judgments in tasks in which it has been alleged that "information about base rates is generally observed to be ignored" (Evans & Bradshaw 1986, p. 16).

Research on the base-rate fallacy is not the only line of research in behavioral decision making where feedback and repetition seem to matter. Another example is research on "preference reversals," in which most participants choose gamble A over B but then state that their minimum willingness-to-accept price for A is less than the price of B (Lichtenstein & Slovic 1971). This basic finding has been replicated many times with a great variety of gambles. In a repeated context, however, preference reversals are not as recalcitrant as this research makes them seem. For instance, Berg et al. (1985), Hamm (reported in Berg et al. 1985), and Chu and Chu (1990) observed that the number of preference reversals decreases if participants repeat the experiment. Berg et al. (1985) concluded that "these findings are consistent with the idea that economic theory describes the asymptotic behavior of individuals after they have become acclimated to the task" (p. 47). Chu and Chu (1990), who embedded their study in a market context, concluded that "*three transactions were all that was needed to wipe out preference reversals completely*" (p. 909, *their emphasis*).

Some have questioned the importance of learning (e.g., Brehmer 1980). Thaler (1987), among others, has argued that the equilibrium and convergence argument is misguided because "when major decisions are involved, most people get too few trials to receive much training" (p. 122). While it may be true that for some situations there is little opportunity for training, it is noteworthy that novices in real-life settings often have the opportunity to seek advice from others in high-stake "first trials" – an option not available in most experiments in both psychology and economics. Moreover, in the first trial, a novice might use a range of other strategies, such as trying to convert the task into hedge trimming rather than tree felling (Connolly 1988) in order to get feedback, holding back reserves, or finding ways to avoid firm commitments.

To conclude this section, testing a stimulus (e.g., a gamble, an inference task, a judgment task, or a choice task) only once is likely to produce high variability in the obtained data (e.g., less consistency in the cognitive processes). In the first trial, the participant might still be in the process of trying to understand the experimental instructions, the setting, the procedure, and the experimenter's intentions. The more often the participant works on the same stimulus, the more stable the stimulus interpretation (and the less pronounced the test anxiety; Beach & Phillips 1967) and the resulting behavior (as long as the situation is incentive-compatible and participants are neither bored

nor distracted). People's performance in early trials, in other words, does not necessarily reflect their reasoning *competence* in later trials. We propose that psychologists consider using stationary replication, that is, repetition of one-shot decisions and game situations as well as feedback, and not restrict their attention to one-shot trials in which participants may be confused and have not had an opportunity to learn.

Last but not least, which design is appropriate is not only a methodological issue. The appropriateness of a design depends crucially on what aspects of behavior and cognition a given theory is designed to capture. Although recently economists have become increasingly interested in learning, prevailing theories in economics still focus on equilibrium behavior. In contrast, many (but not all) psychological judgment and decision-making theories are not explicit about the kind of behavior they target – first impressions, learning, or equilibrium behavior – and also do not explicate how feedback and learning may affect it. Clearly, if theories in psychology were more explicit about the target behavior, then the theories rather than the experimenter would define the appropriate test conditions, and thus questions about whether or not to use repeated trials would be less likely to arise.

4. Financial Incentives versus No Incentives

Although important objections have been raised to the way financial incentives are often structured (e.g., Harrison 1989; 1992), experimental economists who do not use them at all can count on not getting their results published. Camerer and Hogarth (1999) reported that “a search of the *American Economic Review* from 1970–1997 did not turn up a single published experimental study in which subjects were not paid according to performance” (p. 31). As Roth (1995) observed, “the question of actual versus hypothetical choices has become one of the fault lines that have come to distinguish experiments published in the economics journals from those published in psychology journals” (p. 86).

Economists use financial incentives for at least four reasons. The first is the widespread belief among experimental economists that salient payoffs (rewards or punishment) reduce performance variability (Davis & Holt 1993, p. 25). The second is the assumption that the saliency of financial incentives is easier to gauge and implement than most alternative incentives. The third is the assumption that most of us want more of it (so it is fairly reliable across participants), and there is no satiation over the course of an experiment (not so with German chocolate cake, grade points, etc.). The fourth, and arguably the most important argument motivating financial incentives, is that most economics experiments test economic theory, which provides a comparatively unified framework built on maximization assumptions (of utility, profit, revenue, etc.) and defines standards of optimal behavior. Thus, economic theory lends itself to straightforward translations into experiments employing financial incentives.

This framework is sometimes interpreted as exclusively focusing on the monetary structure at the expense of the social structure. We believe this to be a misunderstanding. Every experiment that employs financial incentives implicitly also suggests something about other motivators (e.g., altruism, trust, reciprocity, or fairness). For example, if in prisoner's dilemma games (or public good, trust, ultimatum, or dictator games) the behavior of participants does not correspond to the game-theoretic predictions, that is, if they show more altruism (trust, reciprocity, or fairness) than the theory predicts, then these findings also tell us something about the other nonmonetary motivators (assuming that demand effects are carefully controlled, and the experiments successfully implement the game-theoretic model).

Psychologists typically do not rely on a similarly unified theoretical framework that can be easily translated into experimental design. Moreover, in some important psychological domains, standards of optimal behavior are not as clearly defined (e.g., in mate choice), if they can be defined at all, or conflicting norms have been proposed (e.g., in hypothesis testing, probabilistic reasoning). In addition, there is the belief that "our subjects are the usual middle-class achievement-oriented people who wish to provide [maximal performance]" (Dawes 1996, p. 20), which seems to suggest that financial incentives are superfluous. Along similar lines, Camerer (1995) observed that "psychologists presume subjects are cooperative and intrinsically motivated to perform well" (p. 599).

To quantify how different the conventions in economics and psychology are with regard to financial incentives, we examined all articles published in the *Journal of Behavioral Decision Making (JBDM)* in the 10-year period spanning 1988 (the year the journal was founded) to 1997. We chose *JBDM* because it is one of the major outlets for behavioral decision researchers and provides a reasonably representative sample of the experimental practices in this domain. As our unit of analysis we took experimental studies – a typical *JBDM* article reports several – in which some kind of performance criterion was used, or in which participants were provided with an explicit choice scenario involving monetary consequences.

In addition to studies in which no performance criterion was specified, we excluded studies in which no financial incentives could have been employed because experimenters compared performance across rather than within participants (i.e., between-subjects designs). In addition, we excluded studies in which the main focus was not on the performance criterion – either because it was only one among many explored variables or because processes rather than outcomes were examined. Finally, we omitted studies in which experimenters explicitly instructed participants that there were no right or wrong answers, or that we could not classify unequivocally (e.g., ambiguous performance criteria, or the description of the study leaves it open whether financial incentives were employed at all).

Our criteria were intentionally strict and committed us to evaluating each study in its own right and not with respect to some ideal study (e.g., we did

not assume that each study that explored the understanding of verbal and numerical probabilities could have employed financial incentives only because Olson & Budescu, 1997, thought of an ingenious way to do it). These strict criteria stacked the deck against the claim that psychologists hardly use payments, as studies that could have employed payments if run differently were excluded.

We included 186 studies in the analysis. Out of those 186 studies, 48 (26%) employed financial incentives. Since *JBDM* publishes articles at the intersection of psychology, management sciences, and economics, and experimental economists such as John Hey and David Grether are on the editorial board, this ratio is very likely an overestimate of the use of financial incentives in related domains of psychological research. If one subtracts studies in which at least one of the authors is an economist or is affiliated with an economics department, then the percentage of studies using financial incentives declines to 22% (40 of 178 studies). If one additionally subtracts studies in which at least one of the authors is one of the few psychologists in behavioral decision making who frequently or exclusively use monetary incentives (Budescu, Herrnstein, Rapoport, and Wallsten), then the ratio declines still further to 15% (25 of 163). This survey suggests that financial incentives are indeed not the norm in behavioral decision making.

Our conclusion is also supported by a second sample of studies that we analyzed. As described in section 3, we examined 106 studies on Bayesian reasoning. These studies were published in a variety of journals, including journals from social psychology (e.g., *Journal of Personality and Social Psychology*, *Journal of Experimental Social Psychology*), cognitive psychology (e.g., *Cognition*, *Cognitive Psychology*), and judgment and decision making (e.g., *Organizational Behavior and Human Decision Processes*, *JBDM*). Thus, this sample represents a cross-section of journals. Of these 106 base-rate studies, only three provided financial incentives (Studies 1 and 2 in Nelson et al. 1990; and possibly Kahneman & Tversky's 1973 study).

4.1. Do Financial Incentives Matter?

Given the typical economist's and psychologist's sharply diverging practices, it is not surprising to see diverging answers to the question of whether financial incentives matter. There is overwhelming consensus among economists that financial incentives affect performance for the better (e.g., Davis & Holt 1993; Harrison 1992; Roth 1995; Smith 1991; Smith & Walker 1993a; 1993b). Consequently, experimental economists have debated the "growing body of evidence [from psychology] – mainly of an experimental nature – that has documented systematic departures from the dictates of rational economic behavior" (Hogarth & Reder 1987, p. vii; see e.g., Kahneman et al. 1982; Kahneman & Tversky 1996; Tversky & Kahneman 1981), often on the grounds that such

departures have been shown primarily in experiments without financial incentives (e.g., Smith 1991, p. 887).

The rationale behind this criticism is that economists think of "cognitive effort" as a scarce resource that people have to allocate strategically. If participants are not paid contingent on their performance, economists argue, then they will not invest cognitive effort to avoid making judgment errors, whereas if payoffs are provided that satisfy saliency and dominance requirements (Smith 1976; 1982; see also Harrison 1989 and 1992³), then "subject decisions will move closer to the theorist's optimum and result in a reduction in the variance of decision error" (Smith & Walker 1993a, p. 260; there is an interesting link to the psychology studies on the relationship between "need for cognition" and the quality of decision making: see, e.g., Smith & Levin 1996). Believers in the reality of violations of rational economic behavior in both psychology and economics have dismissed this criticism (e.g., Thaler 1987; Tversky & Kahneman 1987).

Our 10-year sample of empirical studies published in *JBDM* was not selected to demonstrate whether financial incentives matter; therefore it can add systematic empirical evidence. Recall that in our sample of *JBDM* studies, 48 of 186 studies (26%) employed financial incentives. In only 10 of those 48 studies, however, was the effect of payments systematically explored, either by comparing a payment to a nonpayment condition or by comparing different payment schemes. What results were obtained in those 10 studies?

For the studies in which the necessary information was given, we calculated the effect size *eta*, which can be defined as the square root of the proportion of variance accounted for (Rosenthal & Rosnow 1991). *Eta* is identical to the Pearson product-moment correlation coefficient when $df = 1$, as in the case when two conditions are being compared. According to Cohen's (1988) classification of effect sizes, values of *eta* of .1, .3, and .5 constitute a small, medium, and large effect size, respectively. As can be seen in Table 2, the effect sizes for financial incentives ranged from small to (very) large, confirming findings in other review studies (e.g., Camerer & Hogarth 1999).

In the majority of cases where payments made a difference, they improved people's performance. Specifically, payments decreased a framing effect (Levin et al. 1988), made people take the cost of information into account, and increased their confidence in decisions based on highly diagnostic information (Van Wallendaël & Guignard 1992). In an auction experiment, payments brought bids closer to optimality and reduced data variability (Irwin et al. 1992). Payments also decreased the percentage of ties in gamble evaluations relative to nonviolations of the dominance principle (Mellers et al. 1995) and, when combined with "simultaneous" judgment, eliminated preference reversals (Ordóñez et al. 1995). In addition, payments reduced the noncomplementarity of judgments (Yaniv & Schul 1997), brought people's allocation decisions closer to the prescriptions of an optimal model (when self-regarding behavior could be punished; Allison & Messick 1990), and induced people to expend

Table 2: Effects of performance-based payments obtained in studies from *Journal of Behavioral Decision Making*. We calculated effect size η^2 and d (Cohen 1988; Rosenthal & Rosnow 1991) when sufficient information was available

Authors	Task	Comparison	General results (effect size)
Levin, Chapman & Johnson (1988)	Judgments of the likelihood of taking gambles, and expressing confidence in those judgments. Probability information was framed either positively as "percent chance of winning" or negatively as "percent chance of losing"	<p><i>Study 1</i> Hypothetical gambles</p> <p><i>Study 2</i> Gambles with real money; winnings ranged from 25¢ to \$5, averaging about \$2</p>	Reduced framing effect, that is, difference in likelihood of taking the gamble depending on positive vs. negative framing was smaller in real-money gambles than in hypothetical gambles. ^a In both studies, confidence ratings were higher for incomplete than for complete information trials in the explicit inference condition.
Allison & Messick (1990)	Division of shared resource: Participants decide how much they request from a resource, which they share with others	<p><i>Low-payoff condition</i> Resource \$9, \$10.50</p> <p><i>High-payoff condition</i> Resource \$12, \$13.50</p>	No "significant" main effect of payoff condition (p. 201); however, the proportion of the pool requested was on average 24.1% and 29.1% (see Allison & Messick's Table 1) in the low- and high-payoff condition, respectively. Moreover, proportion of requested resource was closest to normative model in the high-payoff condition, in which the pool was nondivisible and other group members could not punish self-regarding behavior (45.8%). ^b
Irwin, McClelland & Schulze (1992)	Gambles in which participants face a 1% chance of a \$40 loss. In an auction procedure, they could enter (optimal) bids for insurance against this loss	<p><i>Hypothetical reward</i> <i>Real reward</i></p> <p>Participants received an initial stake of \$50 or \$80 and kept the positive balance at the conclusion</p>	For hypothetical reward, the bids (in terms of σ) were more variable ($\eta^2 = .60$). There were more very high bids ($\eta^2 = .46$) and more "0" bids ($\eta^2 = .62$) for the insurance in the hypothetical- than in the real-reward condition. In addition, bids were more optimal (measured in terms of the difference between expected value of the insurance and bids) in the real-reward than in the hypothetical-reward condition (difference: $\eta^2 = .39$; absolute difference: $\eta^2 = .45$).

Van Wallendaal & Guignard (1992)	Categorization of stimuli with the aid of information of varying costs and diagnosticity levels ^a	<p><i>Study 1</i> Hypothetical points for correct identification, and hypothetical cost for each piece of information</p> <p><i>Study 2</i> \$1.25 for every correct identification, minus 5¢, 10¢, and 15¢ per piece of information</p>	<p>A main effect of information diagnosticity (inconsistent with the expected-value approach) was larger in Study 2 ($\eta^2 = .79$) than in Study 1 ($\eta^2 = .21$); no "significant" main effect for information costs in Study 1 but a "significant" main effect consistent with the expected-value approach in Study 2 ($\eta^2 = .70$).</p> <p>In both studies, participants expressed greater confidence in decisions based on high-diagnosticity questions than on low-diagnosticity questions; this effect, however, was larger in Study 2 than in Study 1 ($\eta^2 = .24$, $\eta^2 = .74$). In both studies, there were no significant differences in confidence owing to answer diagnosticity or information costs, nor significant interactions.</p> <p>According to an analysis of variance, the effects of incentive on various search measures were not significant (p. 87). In a "partial least square" analysis, the authors found that in the incentive condition "more external search and internal (non-search) processing and . . . significantly more effort overall" (p. 97) was expended.^d</p>
Hulland & Klermmuntz (1994)	Choice of preferred alternative based upon searched information	<p><i>Incentive condition</i> Incentive to spend more effort (one could win a year's subscription to a chosen alternative among women's magazines)</p> <p><i>Control condition</i> Nothing at stake</p> <p><i>Study 2</i> Hypothetical costs for information and wrong decisions, and hypothetical payoffs for correct decisions</p> <p><i>Study 3</i> Real costs for information and real payoff for correct decisions (the highest scorer among 10 scorers wins \$10)</p>	<p>"The monetary prize had virtually no effect on subjects' performance. Results of Experiment 3 replicated all of the major findings of Experiment 2" (p. 259). For instance, participants in both studies overbought in comparison to optimal buys calculated from an expected-value model.^e</p>
Van Wallendaal (1995)	Judgments of the guilt or innocence of suspects in fictional crimes (based on the purchase of information) and probability ratings of those judgments being correct		

(Continued)

Table 2: (Continued)

Authors	Task	Comparison	General results (effect size)
Mellers, Beretty & Birnbaum (1995)	Evaluation of the worth of gambles ^d	Study 2 Hypothetical prices in the baseline condition; in the incentives condition a pair of gambles was randomly selected, and the gamble that was preferred was actually played	"There are no significant main effects or interactions involving financial incentives" (p. 210). However, the mean percentage of violations, nonviolations, and ties across 10 gamble pairs was strongly affected. In the baseline condition 38%, 19%, and 43% were violations, nonviolations, and ties, whereas the proportions in the incentives condition were 36%, 50%, and 16% (Compare Exhibits 7 and 8, similar results for Exhibits 11 and 12). If one only considers violations and nonviolations, then in all 10 gamble pairs in Exhibit 7 (baseline condition), the proportion of violations exceeded that of nonviolations. In contrast, in Exhibit 8 (incentives condition) this is only true for 2 of the 10 gamble pairs.
Ordóñez, Mellers, Chang & Roberts (1995)	Evaluation (attractiveness ratings, strength of preference judgments, selling prices) of pairs of gambles in a sequential vs. simultaneous way (i.e., evaluate pair of gambles on two types of response modes before evaluating the next pair)	No financial Incentives Financial incentives A pair of gambles was randomly selected, and the gamble that was preferred was actually played (payments ranged from \$0 to \$11.34)	Without incentives, only in one of the three combinations of response modes were reversals "significantly" reduced ($\eta^2 = .27$) when comparing sequential to simultaneous performance. With incentives, reversal rates were "significantly" reduced in all three combinations of two response modes ($\eta^2 = .33, .49, .51$) when comparing sequential to simultaneous performance. Ordóñez et al. concluded that "simultaneous performance with financial incentives virtually eliminated reversal rates" (p. 271-72), and "although preference reversals are quite robust, the present results show that subjects can give consistent responses with certain procedures and incentives" (p. 276).
Yaniv & Schul (1997)	Selection of answers from a set of multiple alternatives under two different framings: Include likely alternatives from initial set versus exclude the least likely alternatives	Study 1 No payment Study 2 Payoff rule (which reflected a tradeoff between the number of marked alternatives and probability that the correct choice was marked)	The accuracy of participants' selections is almost identical in Studies 1 and 2 (50% vs. 49%). Although exclusion and inclusion conditions are still significantly different, the difference is reduced in Study 2. The difference in the size of the choice set is larger in Study 1 than in Study 2. Expressed in percentage of the full set, the difference between the inclusion and the exclusion set is 31.5 ($d = 3.2$) percentage points in Study 1 and 20 percentage points ($d = 1.9$) in Study 2. In addition, the difference in accuracy between the inclusion and the exclusion condition is smaller in Study 2 (18 percentage points, $d = 1.2$) than in Study 1 (38 percentage points, $d = 2.7$).

Beeler & Hunton (1997)	Allocation decisions into investment portfolios, and the effect of negative feedback on performance of commitment and information search	No-pay condition and Salary condition Choice of two state lottery tickets or \$2, irrespective of how well their investments performed Contingent compensation Earning of state lottery tickets or cash based on investment performance	As a function of compensation methods, an escalation of commitment ("sunk cost effect") was observed, that is, the average dollars subsequently invested in losing companies was \$35,890 (no-pay), 64,280 (salary), and \$119,450 (contingent). Amount of time participants viewed prospective information decreased and amount of time for retrospective information increased as a function of compensation method (prospective time: 360.0, 258.5, and 151.5 seconds; retrospective time: 218.5, 346.0, and 469.0 seconds). ⁹
------------------------	--	--	--

*Levin et al. concluded that "high levels of personal involvement, such as created by providing real monetary consequences to gambles, can serve to . . . reduce the information framing effect" (p. 39).

^bAs one normative model, Allison and Messick (p. 197) assumed that each participant should have requested all that was permissible, leaving a minimal amount for the final group member. Based on their results, they concluded that "subjects in social decision tasks involving shared resources cannot be modeled as strategic money maximizers" (p. 195). We point out, however, that there is a major methodological problem in this study, namely that of the lack of control of the "social distance" between experimenter and participant. Confronted with persistent deviations from game theoretic predictions in dictator and ultimatum games, Hoffman et al. (1996) manipulated instructional and procedural aspects of the design and implementation of dictator games and found that increasing social distance – in the most extreme case, no one, including the experimenter or any observer of the data, could possibly know any participant's decision (i.e., complete anonymity) – monotonically reduces the deviations from game-theoretic predictions. In Allison and Messick's study, anonymity was guaranteed neither between participant and experimenter nor among participants ("each member would be told what the previous members took," p. 199).

^cAs a normative model, Van Wallendael and Guignard assumed that if participants used an expected-value approach their information purchases should show main effects of cost (with more information being purchased at lower costs) and question diagnosticity (with more high-diagnosticity information being purchased), and other interaction effects. However, they also noted that an alternative model, which better predicted people's information purchases is "not necessarily normatively suboptimal" (p. 36).

^dHolland and Kleinmuntz pointed out that the link between decisions and their consequences may have been small, as many male participants may have been unaffected by the incentive manipulation (subscription to a women's magazine). They suggest that this may explain why incentives did not affect another search measure (participants' use of summary evaluations).

^eWe will not report effect sizes for the multiple statistical comparisons (the study included a $2 \times 2 \times 3$ factorial design and two dependent measures).

^fAs a normative model, Mellers et al. assumed the dominance principle, which holds that increasing one or more outcomes of a gamble should increase the judged price of the gamble, with everything else held constant.

^gThe authors concluded from their results that "performance-based incentives, (i.e., contingent) led to higher escalation of commitment," and that "it appears as though individuals who heightened their commitment to an initial course of action endeavored to resolve cognitive dissonance, justify past decisions, and account for prior actions by searching for retrospective, supporting information" (p. 88). Note, however, that all participants received the false information that the data they studied prior to their investment decisions represented actual cases, and they also were given the feedback, independent of their specific investment decision, that their investment portfolio had declined in value. If participants in the contingent-compensation condition spent more cognitive effort prior to their decision, their search of retrospective investment may not express their increasing commitment but their potential disbelief concerning the feedback fabricated by the experimenter.

more effort (in terms of external search and internal nonsearch processing) in making choices (Holland & Kleinmuntz 1994). In only two cases did payments seem to impair performance: They escalated commitment and time spent obtaining retrospective information (sunk cost effect, Beeler & Hunton 1997; but see the methodological problems mentioned in Table 2, footnote g) and accentuated a (suboptimal) information diagnosticity effect (Van Wallendael & Guignard 1992).

In a few cases, payments did not make a difference. As Table 2 shows, they did not improve either confidence judgments (Levin et al. 1988; Van Wallendael & Guignard 1992) or patterns of information purchase and probability ratings based on that information (Van Wallendael 1995). They also did not decrease the proportion of violations of the dominance principle (Mellers et al. 1995), nor did they increase the accuracy of participants' responses to general knowledge items (Yaniv & Schul 1997).

Given that Table 2 reports *all* studies of the JBDM sample that systematically explored the effect of financial incentives, we conclude that, although payments do not guarantee optimal decisions, in many cases they bring decisions closer to the predictions of the normative models. Moreover, and equally important, they can reduce data variability substantially. These results are in line with Smith and Walker's (1993a) survey of 31 experimental studies reporting on the effects of financial incentives and decision costs (including, e.g., Grether & Plott's 1979 study of preference reversals). Specifically, Smith and Walker (1993a) concluded that "in virtually all cases rewards reduce the variance of the data around the predicted outcome" (p. 245, see further evidence in Grether 1980; Harless & Camerer 1994; Jamal & Sunder 1991).

Aside from the Smith and Walker study, four other recent review articles have explored the effect of financial incentives. First, Camerer and Hogarth (1999) reviewed 74 studies (e.g., on judgment and decision making, games, and market experiments) and compared the behavior of experimental participants who did and did not receive payments according to their performance. Camerer and Hogarth found cases in which financial incentives helped, hurt, did not make a difference, and made a difference although it was not clear whether for better or worse because there was no standard for optimal performance. More specifically, however, Camerer and Hogarth found that financial incentives have the largest effect in "judgment and decision" studies—our focus and running example of the sharply differing practices between experimental economists and psychologists: Out of 28 studies, in 15, financial incentives helped, in 5, they did not have an effect, and in 8, they had negative effects. Regarding the latter, however, Camerer and Hogarth wrote that the "effects are often unclear for various methodological reasons" (p. 21). Moreover, Camerer and Hogarth reported that in many of those studies in which incentives did not affect mean performance, they "*did* reduce variation" (p. 23, their emphasis).

Second, Harrison and Rutstroem (in press), drawing on 40 studies, accumulated overwhelming evidence of a "hypothetical bias" in value elicitation methods. Simply put, they found that when people are asked hypothetically what they would be willing to pay to maintain an environmental good (e.g., the vista of the Grand Canyon), they systematically overstate their true willingness-to-pay (see also Harrison, 1999, for a blunt assessment and methodological discussion of the state of the art of contingent valuation studies). Camerer and Hogarth (1999) mentioned the Harrison and Rutstroem study briefly under the heading "When incentives affect behavior, but there is no performance standard." We believe this to be a misclassification. In our view, true willingness-to-pay is a norm against which "cheap talk" can be measured.

Third, in a meta-analytic review of empirical research (from several applied psychology journals), Jenkins et al. (1998) found financial incentives to be related to performance quantity (e.g., exam completion time) but not quality (e.g., coding accuracy; the authors stressed that this result ought to be "viewed with caution because it is based on only six studies," p. 783). They found an effect size for performance quantity of .34 (point-biserial correlation), which is considered to be of medium size (e.g., Rosenthal & Rosnow 1991). In addition, they reported that the relation between financial incentives and performance is weakest in laboratory experiments (as compared, e.g., to field experiments) – possibly because "laboratory studies typically use small incentives" (Jenkins et al. 1998, p. 784). While their review does not address the impact of financial incentives on intrinsic motivation directly, they concluded that "our results . . . go a long way toward dispelling the myth that financial incentives erode intrinsic motivation" (p. 784). Fourth, and of relevance in light of Jenkins et al.'s results, Prendergast (1999) reviewed the effect of incentive provision in firms and found that there is a positive relationship between financial incentives and performance.

To conclude, concerning the controversial issue of the effects of financial incentives, there seems to be agreement on at least the following points: First, financial incentives matter more in some areas than in others (e.g., see Camerer & Hogarth's 1999, distinction between judgment and decision vs. games and markets). Second, they matter more often than not in those areas that we explore here (in particular, research on judgment and decision making), which are relevant for both psychologists and economists. Third, the obtained effects seemed to be two-fold, namely, convergence of the data toward the performance criterion and reduction of the data's variance. Based on these results, we propose that psychologists in behavioral decision making consider using financial incentives. Although "asking purely hypothetical questions is inexpensive, fast and convenient" (Thaler 1987 p. 120), we conjecture that the benefits of being able to run many studies do not outweigh the costs of generating results of questionable reliability (see also Beattie & Loomes 1997, p. 166).

In addition, only by paying serious attention to financial incentives can psychologists conduct systematic research on many open issues. (For instance, under which conditions do financial incentives improve, not matter to, or impair task performance? For previous research on these conditions, see, e.g., Beattie & Loomes 1997; Hogarth et al. 1991; Payne et al. 1992; Pelham & Neter 1995; Schwartz 1982; Wilcox 1993.)⁶ How do incentives (and opportunity costs) affect decision strategies and information processing (e.g., Payne et al. 1996; Stone & Schkade 1994; Wallsten & Barton 1982;), and how do they interact with other kinds of incentives (e.g., social incentives) and motives?⁷ Some of the reported research also highlights the need to understand better how incentives interact with other variables of experimental design (e.g., repetition of trials, Chu & Chu 1990, and presentation of gambles, Ordóñez et al. 1995; see also Camerer 1995, sect. I and Camerer & Hogarth 1999), and to establish what kinds of salient and dominant rewards are effective (e.g., the problem of flat maxima, see Harrison 1994; von Winterfeldt & Edwards 1982).

Ultimately, the debate over financial incentives is also an expression of the precision of theories or a lack thereof. Economists virtually always pay because the explicit domain of economic theories is extrinsically motivated economic behavior. Psychological theories in behavioral decision making often do not make it completely clear what behavior they target – intrinsically or extrinsically motivated behavior. If theories were more explicit about their domain and the implicated motivation, then they rather than the experimenters would define the appropriate test conditions.

We conclude this section by briefly discussing two possible reasons for mixed results (sect. 4.1.1), and whether and how payments affect intrinsic motivation (sect. 4.1.2).

4.1.1. *Reasons for the Mixed Results?*

The majority of the results in Table 2 are inconsistent with studies that did not find any effect of payment (see, e.g., the studies mentioned in Dawes 1988; Hogarth et al. 1991; Stone & Ziebart 1995). How can these discrepant results be explained? There are at least two possible explanations. The first was pointed out by Harrison (1994, p. 240), who reexamined some of Kahneman and Tversky's studies on cognitive illusions that used financial incentives and concluded that the majority of these experiments lack payoff dominance (see note 5). In other words, not choosing the theoretically optimal alternative costs participants in these experiments too little. Based on new experiments (e.g., on preference reversals and base-rate neglect) that were designed to satisfy the dominance requirement, Harrison (1994) concluded that in his redesigned experiments, observed choice behavior is consistent with the predictions of economic theory.

A second possible explanation can be drawn from the existence of multiple and contradictory norms against which performance might be compared

(see, e.g., the controversy between Kahneman & Tversky 1996 and Gigerenzer 1996; see also Hilton 1995 on the issue of conversational logic). The problem of multiple and ambiguous norms may be compounded by a focus on coherence criteria (e.g., logical consistency, rules of probability) over correspondence criteria, which relate human performance to success in the real world (e.g., speed, accuracy, frugality). Clearly, if multiple norms exist and the experimenter does not clarify the criterion for which participants should aim (e.g., by specification of payoffs), then payment will not necessarily bring their responses closer to the normative criterion the experimenter has in mind. More generally, as argued by Edwards (1961):

Experiments should be designed so that each subject has enough information to resolve ambiguities about how to evaluate the consequences of his own behavior which are inherent in conflicting value dimensions. That means that the subject should have the information about costs and payoffs . . . necessary to evaluate each course of action relative to all others available to him. (p. 283)

4.1.2. *How Do Financial Incentives Affect Intrinsic Motivation?*

An important argument against the use of financial incentives is that they crowd out intrinsic motivation (if it exists). This argument can be traced back to Lepper et al.'s (1973) finding that after being paid to perform an activity they seemed to enjoy, participants invested less effort in the activity when payoffs ceased. Lepper et al. interpreted participants' initial apparent enjoyment of the activity as evidence of intrinsic motivation and their subsequent decrease in effort expenditure as evidence of the negative impact of extrinsic rewards on intrinsic motivation. A huge literature has evolved consequently. Drawing on an extensive meta-analysis by Cameron and Pierce (1994), Eisenberger and Cameron (1996) performed a meta-analysis on the question of whether financial incentives really undermine intrinsic motivation.

Based on their examination of two main measures of intrinsic motivation, namely, the free time spent on the task post-reward and the expressed attitude toward the task, they did not find that completion-dependent reward (i.e., reward for completing a task or solving a problem) had any negative effect. Moreover, they found that quality-dependent reward (i.e., reward for the quality of one's performance relative to some normative standard) had a positive effect on expressed attitudes toward the task. Ironically, the only measure on which Eisenberger and Cameron (1996) found a reliable negative effect was the free time spent carrying out the activity following performance-independent reward (i.e., reward for simply taking part in an activity), the type of reward commonly used in psychological experiments. Eisenberger and Cameron (1996) concluded that "claimed negative effects of reward on task interest and creativity have attained the status of myth, taken for granted despite

considerable evidence that the conditions producing these effects are limited and easily remedied" (p. 1154).

The conclusions of Cameron and colleagues have been challenged (e.g., Deci et al. 1999a; Kohn 1996; Lepper et al. 1996; see also the debate in the *American Psychologist*, June 1998). In their most recent meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation, Deci et al. also discussed the procedure employed by Eisenberger and Cameron (1996). No surprise, they come to very different conclusions, confirming the classic finding that tangible rewards (i.e., financial incentives) undermine intrinsic motivation. One important bone of contention is the definition of the relevant set of studies. Deci et al. argue that it ought to be confined to "interesting" tasks, and ought to exclude "boring" tasks, some of which Eisenberger and Cameron include (see also Deci et al. 1999b; Eisenberger et al. 1999; Lepper et al. 1999). In sum, there is agreement that rewards can be used as a technique of control; disagreement exists as to unintended consequences of rewards. We believe that the situation calls for a meta-analysis done by the two camps and a jointly determined arbiter following the model of "adversarial collaboration" proposed by Kahneman and exemplified in Mellers et al. (2001). In the meantime, we believe that the boring nature of many experiments and the available evidence reported here suggest that financial incentives matter in tasks examined in behavioral decision making (see Table 1; Camerer & Hogarth 1999) and thus ought to be considered, unless previous studies show that financial incentives do not matter for a particular task.⁸

5. Honesty versus Deception

Deceiving participants is generally taboo among experimental economists (Davis & Holt 1993, p. 24) and, indeed, economics studies that use deception can probably be counted on two hands.⁹ Davis and Holt (1993, pp. 23–24; see also, Hey 1991; Ledyard 1995) gave the following typical rationale for economists' reasons to argue against deception (for a rare dissenting view in economics, see Bonetti 1998, but see also the comments of Hey 1998; McDaniel & Starmer 1998):

The researcher should ... be careful to avoid deceiving participants. Most economists are very concerned about developing and maintaining a reputation among the student population for honesty in order to ensure that subject actions are motivated by the induced monetary rewards rather than by psychological reactions to suspected manipulation. Subjects may suspect deception if it is present. Moreover, even if subjects fail to detect deception within a session, it may jeopardize future experiments if the subjects ever find out that they were deceived and report this information to their friends.

Even if participants initially were to take part in experiments out of a sense of cooperation, intrinsic motivation, or the like, economists reason that they

will probably become distrustful and start second-guessing the purpose of experiments as soon as they hear about such deception. In other words, economists fear reputational spillover effects of deceptive practices even if only a few of their tribe practice it. In the parlance of economists, participants' expectation that they will not be deceived (i.e., honesty on the part of the experimenter) is a common good of sorts (such as air or water) that would be depleted (contaminated) quickly if deception was allowed and the decision about its use left to each experimenter's own cost-benefit analysis. On theoretical and empirical grounds, economists do not trust experimenters to make an unbiased analysis of the (private) benefits of deception and its (public) costs. The temptation, or, in economists' parlance, the "moral hazard" to capture the private benefits of deception is perceived to be simply too strong. Indeed, given that the American Psychological Association (APA) ethics guidelines (APA 1992, p. 1609) propose to employ deception as a last-resort strategy, to be used only after careful weighing of benefits and costs, the frequent use of deception in some areas of psychology seems to confirm economists' fear.

Take the highest ranked journal in social psychology, the *Journal of Personality and Social Psychology* (*JPSP*), and its predecessor, *Journal of Abnormal and Social Psychology*, as an illustration. After a sharp upswing during the 1960s (where it tripled from 16% in 1961 to 47% in 1971), the use of deception continued to increase through the 1970s, reaching its high in 1979 (59%) before dropping to 50% in 1983 (Adair et al. 1985). Since then it has fluctuated between 31% and 47% (1986: 32%, 1992: 47%, 1994: 31%, 1996: 42%; as reported in Epley & Huff 1998; Nicks et al. 1997; Sieber et al. 1995).

While some of these fluctuations may reflect different definitions of what constitutes deception (e.g., compare the more inclusive criteria employed by Sieber et al. with the criteria used by Nicks et al.), a conservative estimate would be that every third study published in *JPSP* in the 1990s employed deception. (In other social psychological journals, e.g., *Journal of Experimental Social Psychology*, the proportion is even higher; Adair et al. 1985; Nicks et al. 1997.) The widespread use of deception in social psychology in recent years contrasts markedly with its decidedly more selective use in the 1950s and earlier (Adair et al. 1985). Although deception is likely to be most frequent in social psychology, it is not restricted to it (see sects. 6.1 and 6.3 in the discussion).

Why do psychologists use deception? Although some critics of the frequent use of deception attributed it to a "fun-and-games approach" (Ring 1967, p. 117) to psychological experimentation, today's primary motivation for deception seems to rest on at least two serious methodological arguments: First, if participants were aware of the true purpose of a study, they might respond strategically and the investigator might lose experimental control. For instance, one might expect participants to "bend over backwards" (Kimmel 1996, p. 68) to show how accepting they are of members of other races if they know that they are participating in a study of racial prejudices. To the extent that psychologists, more than economists, are interested in social behavior and "sensitive" issues, in

which knowledge of the true purpose of a study could affect participants' behavior (e.g., attitudes and opinions), one might expect deception to be used more often in psychology. The second argument is that deception can be used to produce situations of special interest that are unlikely to arise naturally (e.g., an emergency situation in which bystander effects can be studied).

Despite "widespread agreement" that deception is "often a methodological necessity" (Kimmel 1996, p. 68), and the claim that there is no reason to worry about the methodological consequences of deception (e.g., Christensen 1988; Sharpe et al. 1992; Smith & Richardson 1983), its use has been a longstanding and persistent concern in psychology. Anticipating economists' common good argument, Wallsten (1982) suggested that the erosion of participants' trust would hurt everyone who relies on the participant pool. While some authors proposed cosmetic changes in the use of deception (e.g., Taylor & Shepperd 1996), others proposed more drastic measures (e.g., Baumrind 1985; Kelman 1967; MacCoun & Kerr 1987; Newberry 1973; Ortmann & Hertwig 1997; 1998; Schultz 1969; Vinacke 1954).

5.1. Does Deception Matter?

Our concern here is pragmatic not ethical (see Baumrind 1964; 1971; 1985), that is, we are interested in the methodological consequences of the use of deception on participants' attitudes, expectations, and in particular, on participants' behavior in experiments. Before we discuss the available evidence, it is useful to conceptualize the interaction between participant and experimenter as a one-sided prisoner's dilemma, or principal-agent game. Such a game models the relationship between an agent and a principal, both of whom can either contribute their respective assets (trust for the principal, honesty for the agent) or withhold them. In the current context, the experimenter (agent) can choose either to deceive participants or to be truthful about the setting and purpose of the experiment, while the participant (principal) can choose either to trust the experimenter or to doubt the experimenter's claims. The game-theoretic predictions for a one-shot principal-agent game are, dependent on the parameterization, clear-cut: The agent will defect – at least with some probability. The principal, anticipating the defection, will doubt the experimenter's claims – at least with some probability (see Ortmann & Colander 1997, for two typical parameterizations).

The interaction between agent and principal, of course, is not likely to be a one-shot game. Participants (principals) may come into the laboratory either inexperienced or experienced (by way of previous participation in deception experiments). If they are experienced, then that experience may bear directly on their expectation of the experimenter's action choice. If they are inexperienced, then other participants' experience may still bear on their expectation. If participants have reason to trust the experimenter, they may act like the "good" (Orne 1962) or "obedient" (Fillenbaum 1966) participants they are

often assumed to be in psychology (see Rosenthal & Rosnow 1991). If they have reason to believe that the agent will deceive them, however, their behavior may range from suspicious to apathetic (Newberry 1973) and negativistic (Christensen 1977; Weber & Cook 1972).

Experimental results from trust games suggest that people (participants) may accept being fooled once, but not twice (Dickhaut et al. 1995). Recent results reported by Krupat and Garonzik (1994) also suggest that prior experience with deception affects participants' expectations, that is, increases their suspicion (see also Epley & Huff 1998). According to Krupat and Garonzik (1994), such suspicion is likely to introduce "considerable random noise" into their responses (p. 219). In this context it is interesting to note that Stang (1976) already pointed out that the percentage of suspicious participants (in conformity experiments) tracked closely the increase of the use of deception through the 1960s.

Ironically, the APA ethical guidelines concerning debriefing may exacerbate rather than diminish participants' suspicion: "Deception that is an integral feature of the design and conduct of an experiment must be explained to participants as early as it is feasible, preferably at the conclusion of their participation, but no later than at the conclusion of the research" (APA 1992, p. 1609). From an ethical point of view, debriefing is the right thing to do; from a pragmatic point of view, however, it only undermines the trust of actual and potential participants and thereby contaminates the data collected in future experiments: "Each time this quite proper moral requirement is met, the general impression that psychologists commonly deceive is strengthened" (Mixon 1972, p. 145).

Notwithstanding this concern regarding the use of deception, a number of researchers in psychology have advocated its use on the grounds that participants have a favorable attitude toward it. Smith and Richardson (1983), for example, observed that participants in experiments involving deception reported having enjoyed, and indeed having benefited from, the experience more than those in experiments without deception. Summing up his review of research on the impact of deception on participants, Christensen (1988) concluded: "This review . . . has consistently revealed that research participants do not perceive that they are harmed and do not seem to mind being misled. In fact, evidence exists suggesting that deception experiments are more enjoyable and beneficial than nondeception experiments" (p. 668). In Christensen's (1988) view, "the scale seems to be tilted in favor of continuing the use of deception in psychological research" (p. 664; see also Aitkenhead & Dordoy 1985; Sharpe et al. 1992).

However, even if undergraduate participants tell experimenters (often their professors) the truth about how they feel about deception and genuinely do not mind it (Smith & Richardson 1983), which is by no means a universal finding (e.g., Allen 1983; Cook et al. 1970; Epstein et al. 1973; Fisher & Fyrberg 1994; Rubin 1985; Oliansky 1991), we believe that studies of feelings

about and attitudes toward deception overlook a key issue, namely, the extent to which deception affects participants' *behavior* in experiments. Some intriguing findings suggest that, ironically, it is sometimes the experimenter who is duped in an experiment employing deception. For example, Newberry (1973) found that a high percentage of participants, given a tip-off by an experimental confederate, do not admit to having had foreknowledge when questioned later (30–80% in various conditions) – a result that surely undermines the frequent assumption that participants are cooperative (e.g., Bröder 1998; Kimmel 1998).

MacCoun and Kerr (1987) gave a particularly dramatic example that indicates that participants' behavior is affected by the expectation of deception: When a participant had an epileptic seizure during an experiment, the other participants present appeared to believe the seizure was a charade perpetrated by the experimenter and a confederate and therefore initially ignored it. The only person who immediately helped the victim was the only one who had no prior psychology coursework (MacCoun & Kerr 1987). Along the same lines, Taylor and Shepperd (1996) conducted an experiment in which they used deception to study the effectiveness of conventional debriefing procedures in detecting suspicion of deception. Despite explicit instruction not to communicate while the experimenter left the room on a pretext, participants talked during the experimenter's absence and thereby found out that they were being deceived. In a debriefing, none of them revealed this discovery.

To conclude, because psychology students are the main data source in psychological studies (Sieber & Saks 1989), a substantial proportion of participants can be expected to have experienced deception directly. Owing to students' general expectations (due to coursework) or direct personal experiences, deception can have (negative) consequences even in those domains of psychology in which deception is not or is less frequently used. We therefore concur with the argument advanced by economists and (some) psychologists that participants' trust is a public good worth investing in to increase experimental control. We propose that psychologists view the use of deception as involving a trade-off not only "between methodological and ethical considerations" (Kimmel 1996, p. 71), but also between its methodological costs and benefits.

6. General Discussion

In this article, we have been concerned with practices of psychological experimentation and their divergence from those of experimental economics. In particular, we considered four key variables of experimental design that take on markedly different realizations in the two disciplines. We argued that the conventions in economics of providing and having participants enact a

script, repeating trials, giving financial incentives, and not deceiving participants are *de facto* regulatory, allowing for comparatively little variation in experimental practices between researchers. The corresponding experimental practices in psychology, by contrast, are not regulated by strong conventions. This laissez-faire approach allows for a wide range of experimental practices, which in turn may increase variability in the data obtained and ultimately may impede theoretical advances.

Are our findings consonant with psychologists' and economists' perceptions of their own and the other discipline's practices? Why do we see different realizations of key variables across different disciplines and what are the policy implications of our arguments? In the next sections, we address each of these questions in turn.

6.1. How Researchers Describe Their Own Practices and Those of the Other Discipline

We have provided various illustrations for the two theses we proposed, namely, that (1) key variables of experimental design tend to be realized differently in economics and psychology and (2) experimental standards in economics are regulatory in that they allow for little variation between the experimental practices of individual researchers, whereas experimental standards in psychology are comparatively laissez-faire.

Are these two theses also reflected in the way experimentalists in both fields describe their own practices? We conducted a small-scale survey in which we asked researchers in the fields of behavioral decision making and experimental economics to respond to nine questions concerning the use of financial incentives, trial-by-trial feedback, and deception. The questions asked researchers to describe their own research practices (e.g., "How often do you use performance-contingent payments in your experiments?"), research practices in their field generally (e.g., "How often do you think that experimenters in economics/JDM research use performance-contingent payments?"), and research practice in the related field (e.g., "How often do you think that experimental economists/psychologists use performance-contingent payment?"). Researchers were asked to provide their responses in terms of absolute frequencies ("In ___ out of 10 experiments?"); alternatively, they could mark an "I don't know" option.

We sent the questionnaire to the electronic mailing lists of the *European Association for Decision Making* and the *Brunswik Society*. Both societies encompass mostly European and American psychologists interested in judgment and decision making. We also distributed the questionnaire at the 1999 annual meeting of the *Economic Science Association*, which is attended by experimental economists. A total of 26 researchers in psychology and 40 researchers in economics responded. Admittedly, the response rate for psychologists was

quite low (the response rate for economists was about 60%); both samples, however, encompassed well-established as well as young researchers.

Economists estimated that, on average, they used financial incentives in 9.7 out of 10 experiments ($MD = 10$, $SD = .8$); trial-by-trial feedback in 8.7 out of 10 experiments ($MD = 9$, $SD = 2.1$), and deception in .17 out of 10 experiments ($MD = 0$, $SD = .44$). In contrast, psychologists' average estimates were 2.9 for financial incentives ($MD = 1$, $SD = 3.5$), 2.4 for trial-by-trial feedback ($MD = 1$, $SD = 3.2$), and 1.7 for deception ($MD = 0$, $SD = 2.8$). Aside from the drastically different self-reported practices across fields, the results also demonstrate the wider range of practices within psychology. Concerning financial incentives, for instance, 40% of psychologists responded that they never use financial incentives, whereas 32% use them in half or more of their experiments. Regarding deception, 60% stated that they never use it, whereas 20% use it in half or more of their experiments. When we asked researchers to characterize the general practices in their own field on the same measures, we obtained responses close to those described above. However, researchers in both groups believed that they use financial incentives and trial-by-trial feedback slightly more often and deception slightly less often than researchers in their field as a whole.

To what extent are psychologists and economists aware that experimental practices are different in the other field? Although the psychologists were aware that practices in economics differ from those in their own field, they underestimated the extent of the differences. On average, they estimated that economists use financial incentives in 5.6 out of 10 experiments, give trial-by-trial feedback in 3.2 out of 10 experiments, and use deception in 1.2 out of 10 experiments. Although economists' estimates of the use of financial incentives by psychologists was fairly accurately calibrated ($M = 2.3$), they overestimated the use of trial-by-trial feedback ($M = 4.5$) and deception ($M = 5.5$) by psychologists.¹⁰

The results of our small-scale survey are consistent with the two theses we proposed: Experimental practices in behavioral decision making and economics differ and the research practices of psychologists are much more variable. Although some of this variability is likely to be driven by behavioral decision making researchers' interest in questions that do not lend themselves to the use of financial incentives or trial-by-trial feedback, we suggest that the large variance in their responses also reflects the lack of standards committing them to consistency in experimental practices.

6.2. Why Do the Methodological Practices Differ?

There is no simple answer to this question. Differences in experimental practices are neither recent nor confined to cross-disciplinary comparisons. Danziger (1990) identified at least three diverging models of investigative

practice in early modern psychology: the Wundtian, the clinical, and the Galtonian. According to Danziger, the investigators' different research goals drove different practices. Whether one wanted to learn about pathological states (French investigators of hypnosis), individual differences (Galton), or elementary processes in the generalized human mind (Wundt) determined what investigative situations seemed appropriate. Researchers in contemporary psychology pursue a multitude of research goals as well, and not only those of early modern psychology. To the extent that Danziger's (1990) thesis that different goals give rise to different investigative practices is valid, the heterogeneity of experimental practices within psychology therefore should not be surprising.¹¹

In contrast to psychology, experimental economics displays much less variability in research goals. Roth (1995) identified tests of models of individual choice and game theory (especially those involving industrial organization topics) as the early preoccupations of experimental economists. The later game-theoretic reframing, over the past dozen years, of nearly every field in economics – from microeconomic and industrial organization theory (e.g., Kreps 1990; Tirole 1988) to macroeconomic policy issues (Barro 1990) – provided a unifying theoretical framework that could easily be translated into experimental design.

Yet another aspect that helped to promote the comparative homogeneity of experimental practices within economics was its status as the “new kid on a hostile block” (Lopes 1994, p. 218). In light of severe criticisms from prominent economists who claimed that it was impossible to make scientific progress by conducting experiments (e.g., Lipsey 1979; Russell & Wilkinson 1979; see *The Economist* May 8, 1999, p. 84), it is not surprising that economics was “more self-conscious about its science” (Lopes 1994, p. 218) and methodology than psychology. This explanation suggests that widely shared research goals and the prevalent rational-actor paradigm forced certain conventions and practices on experimental economists in a bid to gain acceptance within their profession. Last but not least it is noteworthy that throughout the 1970s and 1980s, experimental economics was concentrated at about a half dozen sites in the United States and Europe. We conjecture that this concentration helped the comparatively small number of experimental economists to agree on generally accepted rules of experimentation.

To conclude, several factors may account for the differing experimental practices in psychology and economics. Multiple research goals and the lack of a unifying theoretical framework that easily translates into experimental design may have promoted methodological variability in psychology. In contrast, the necessity to justify their practices within the discipline, an unusual concentration of key players in a few laboratories during the take-off phase, and the unifying framework provided by game theory may have helped economists to standardize their methodology.

6.3. Policy Implication: Subject Experimental Practices to Experimentation

As recently argued by Zwick et al. (1999, p. 6), methodological differences between psychology and economics are (at least partly) “derivatives” of differences in the assumptions commonly invoked (explicitly or implicitly) by economists and psychologists in the study of human choice. In our view, this argument must not be read as a justification to do business as usual. Neither psychologists nor economists have reason to avoid an interdisciplinary dialogue on the diverging methodologies for several reasons. First, some of the methodological differences – in particular, the (non)use of deception and scripts, but also the issue of abstract versus “natural” scripts (see note 4) – are not derivatives of theory differences; rather, they seem to be driven by methodological concerns that are largely independent of differences in theories (e.g., trust of potential participants).

Second, even those experimental practices that can be plausibly considered derivatives – for instance, financial incentives and repetition – can also be justified on the grounds of arguments not tightly linked with theory. For instance, it seems widely accepted that financial incentives reduce data variability (increase effect sizes and power of statistical tests; e.g., Camerer & Hogarth 1999; Smith & Walker 1993a). Similarly, a likely benefit of repetition is that participants have the chance to familiarize themselves with all the wrinkles of the unusual situation, and thus, their responses are likely to be more reliable (Binmore 1994).

Third, even if many psychologists do not endorse standard economic theory, they are often (particularly in recent decades) interested in testing its various assumptions (e.g., transitivity of preferences) or predictions. Those tests inevitably do entail the question of what is a “fair” test of standard economic theory – a question to which both psychologists and economists have to find a common answer. Finally, as economists move closer to psychologists’ view of human choice – for instance, Simon’s (1957) notion of bounded rationality, Selten’s (1998) aspiration-adaptation theory, Roth and Erev’s (1995) work on the role of reinforcement learning in games, Camerer and Ho’s (1999) work on reinforcement and belief learning in games, Goeree and Holt’s (in press a; in press b) incorporation of stochastic elements into game theory (see Rabin 1998, for many more examples) – one may envision a long-run convergence toward a common core of axioms in economics and psychology. A common ground concerning methodological practices – based upon an interdisciplinary dialogue and empirically informed design decisions – is likely to promote a theoretical convergence.

How can economists and psychologists establish such a common ground? As we pointed out earlier, we do not hold the conventions and practices in experimental economics to be the gold standard; they bring both benefits and costs. Nevertheless, there is a striking difference between the methodological

approaches in psychology and economics: Economists seem to engage more often in cost-benefit analyses of methodological practices and to be more willing to enforce standards (e.g., to prohibit deception) if they are convinced that their benefits outweigh their costs. We suggest that psychologists, particularly in the context of justification, should also engage more frequently in such cost-benefit analyses and, as researchers, collaborators, and reviewers, enforce standards that are agreed upon as preferable. This is not to say that psychologists should adopt economists' practices lock, stock, and barrel. Rather, we advocate the subjection of methodological practices to systematic empirical (as well as theoretical) analysis. Applied to the variable of financial incentives, such an approach might be realized as follows (see also Camerer & Hogarth 1999).

Researchers seeking maximal performance ought to make a decision about appropriate incentives. This decision should be informed by the evidence available. If there is evidence in past research that incentives affect behavior meaningfully in a task identical to or similar to the one under consideration, then financial (or possibly other) incentives should be employed. If previous studies show that financial incentives do not matter, then not employing incentives can be justified on the basis of this evidence. In cases where there is no or only mixed evidence, we propose that researchers employ a simple "do-it-both-ways" rule. That is, we propose that the different realizations of the key variables discussed here, such as the use or non-use of financial incentives (or the use of different financial incentive schemes), be accorded the status of independent variables in the experiments. We agree with Camerer and Hogarth's (1999) argument that this practice would rapidly give rise to a database that would eventually enable experimenters from both fields to make data-driven decisions about how to realize key variables of experimental design.

This conditional do-it-both-ways policy should also be applied to two other variables of experimental design discussed here, namely, scripts and repetition of trials. In contrast, we propose that the *default* practice should be not to deceive participants, and individual experimenters should be required to justify the methodological necessity of each instance of deception to institutional review boards, referees, and editors. We do not exclude the possibility that there are important research questions for which deception is truly unavoidable.

Nevertheless, we advocate a multi-method approach in which deception is replaced as much as possible by a collection of other procedures, including anonymity (which may undo social desirability effects; see the recent discussion on so-called double-blind treatments in research on dictator games, Hoffman et al. 1996), simulations (Kimmel 1996, pp. 108-13), and role playing (Kimmel 1996, pp. 113-16). We are aware that each of these methods has been or can be criticized (for a review of key arguments see Kimmel 1996). Moreover, it has been repeatedly pointed out that more research is needed

to evaluate the merits of these alternatives (e.g., Diener & Crandall 1978; Kimmel 1996). A do-it-both-ways rule could be used to explore alternatives to deception by comparing the results obtained from previous deception studies to those obtained in alternative designs.

Let us conclude with two remarks on the APA rule of conduct concerning deception:

Psychologists do not conduct a study involving deception unless they have determined that the use of deceptive techniques is justified by the study's prospective scientific, educational, or applied value and that equally effective alternative procedures that do not use deception are not feasible. (APA 1992, p. 1609)

Systematic search for alternative procedures – if enforced – may prove to be a powerful tool for reducing the use of deception in psychology. For instance, of the ten studies reported in Table 2, three used deception (Allison & Messick 1990, p. 200; Beeler & Hunton 1997, p. 83; Irwin et al. 1992, p. 111), including incorrect performance feedback, wrong claims about performance-contingent payments, and rigged randomization procedure. In our view, each of these deceptive practices was avoidable. Deception was also avoidable in another set of studies we reported here. In our sample of Bayesian reasoning studies (see sect. 3), we found that 37 out of 106 (35%) employed some form of deception (e.g., lying to participants about the nature of the materials used, falsely asserting that sampling was random, a precondition for the application of Bayes's theorem). If researchers met the APA requirement to seek alternatives to deception, they would have discovered "equally effective alternative procedures" already in the literature. Research in both psychology (e.g., Wallsten 1972; 1976) and economics (e.g., Grether 1980) shows that one can do without deception completely in research on Bayesian reasoning.

Finally, we propose (in concurrence with a suggestion made by Thomas Wallsten) that the assessment of the "prospective scientific value" of a study should not depend on whether or not a particular study can be conducted or a particular topic investigated. Rather, the question ought to be whether or not a theory under consideration can be investigated without the use of deception. This way, our assessment of the "prospective scientific value" of deception is closely linked to theoretical progress rather than to the feasibility of a particular study.

7. Conclusion

Some of the most serious (self-)criticism of psychology has been triggered by its cycles of conflicting results and conclusions, or more generally, its lack of cumulative progress relative to other sciences. For instance, at the end of the 1970s, Meehl (1978) famously lamented:

It is simply a sad fact that in soft psychology theories rise and decline, come and go, more as a function of baffled boredom than anything else; and the enterprise shows a disturbing absence of that *cumulative* character that is so impressive in disciplines like astronomy, molecular biology, and genetics. (p. 807)

Since the 1970s, psychology's self-esteem has improved – with good reason. For instance, thanks to the increasing use of meta-analytic methods (Glass et al. 1981; Hedges & Olkin 1985), it has become clear that psychology's research findings are not as internally conflicted as once thought. As a result of this, some researchers in psychology have already called off the alarm (Hunter & Schmidt 1990; Schmidt 1992).

Despite this optimism, results in the “softer, wilder areas of our field,” which, according to Rosenthal (1990, p. 775), include clinical, developmental, social, and parts of cognitive psychology, still seem “ephemeral and unreplicable” (p. 775). In his classic works on the statistical power of studies, Cohen (1962; 1988) pointed out two reasons (among others) why this is so. First, in an analysis of the 1960 volume of the *Journal of Abnormal and Social Psychology*, Cohen (1962) showed that if one assumes a medium effect size (corresponding to the Pearson correlation of .40), then experiments were designed in such a way that the researcher had less than a 50% chance of obtaining a significant result if there was a real effect (for more recent analyses, see Rossi 1990; Sedlmeier & Gigerenzer 1989). Second, Cohen (1988) suggested that many effects sought in various research areas in psychology are likely to be small. Whether or not one agrees with this assessment, the important point is that “effects are appraised against a background of random variation” (p. 13). Thus, “the control of various sources of variation through the use of improved research designs serves to increase effect size” (p. 13) and, for that matter, the power of statistical tests as well.

We believe that the realizations of the four key variables of experimental design in the areas of research discussed here contribute to the variability of empirical findings. Based on the evidence reviewed here, we argue that the practices of not providing a precisely defined script for participants to enact, not repeating experimental trials, and paying participants either a flat fee or granting a fixed amount of course credit only leave participants uncertain about the demand characteristics of the social situation “experiment.” The fact that psychologists are (in)famous for deceiving participants is likely to magnify participants' uncertainty and second-guessing.

If our claim that a *laissez-faire* approach to experimentation invites lack of procedural regularity and variability of empirical findings is valid, and the resulting conflicting data indeed strangle theoretical advances at their roots (Loftus, in Bower 1997, p. 356), then discussion of the methodological issues addressed here promises high payoffs. We hope that this article will spur psychologists and economists to join in a spirited discussion of the benefits and costs of current experimental practices.

Notes

1. Sieber and Saks (1989) reported responses of 326 psychology departments. They found that of the 74% that reported having a participant pool, 93% recruited from introductory courses. The authors also found that "only 11% of departments have a subject pool that is voluntary in the strictest sense" (p. 1057). In contrast, economists recruit their participants in more or less randomly determined classes, through flyers or e-mail, often drawing on students from other disciplines. Because economists also typically use financial incentives, it is probably safe to assume that participation is voluntary.
2. For obvious reasons, we cannot reproduce the extensive instructions to participants here. However, we urge the reader who has not yet encountered a script-based study to take a look (e.g., pp. 1247 through 1253 in Camerer et al. 1989).
3. Most of the word problems listed here (e.g., conjunction task, engineer-lawyer task) are classic problems studied in the heuristics-and-biases program. Results and conclusions from this program have been hotly debated (for the different points of view, see the debate between Kahneman & Tversky 1996, and Gigerenzer 1996).
4. Scripts may be content-free or enriched with social context. In an attempt to control home-grown priors (i.e., beliefs and attitudes that participants bring to the experiment), the scripts provided by economists are typically as content-free as possible. From the perspective of the experimenter, such environments may be precisely defined, but they seem to tax the cognitive abilities of participants more than seemingly more complex but familiar real-world scripts, because they take away the "natural" cues that allow participants in real-world environments to understand situations. Assuming the existence of domain-specific reasoning modules, Cosmides and Tooby (1996) even argue that the starkness of laboratory environments prevents specialized inference engines from being activated, and that mismatches between cues and problem types are far more likely under artificial experimental conditions than under natural conditions. This trade-off between control of home-grown priors and accessibility of "natural" cues has long been discussed in psychology (e.g., Bruce 1985; Koriat & Goldsmith 1996 for the real-life/laboratory controversy in memory research; see Goldstein & Weber 1997 for the issue of domain specificity in decision making, and Winkler & Murphy 1973 for their critique of the bookbag-and-poker chips problem in research on Bayesian reasoning). It has also recently been addressed in studies by economists (e.g., Dyer & Kagel 1996; Schotter et al. 1996).
5. Harrison (1989; 1992) argued that many experiments in economics that provide financial incentives dependent on performance nevertheless lack "payoff dominance." Lack of payoff dominance describes essentially flat maxima, which make it relatively inexpensive for participants not to choose the theoretically optimal action (von Winterfeldt & Edwards 1982). The implication of Harrison's critique is that performance in a task can only be classified as "irrational," "inconsistent," or "bounded" if the difference between the payoff for participants' actual behavior and that for optimal behavior in an experiment is monetarily significant to participants given their standard hourly wage. "Significant" could mean, for example, that the potential payoff lost owing to nonoptimal behavior in a one-hour experiment exceeds one hour's worth of wages for the participant and 25% of total payoffs obtainable. If the difference between the payoff for the participant's actual behavior and that for optimal behavior is, say, only 5%, one could argue that the payoff decrement participants accept by not behaving optimally is too trivial to be considered "irrational."
6. The systematic study of financial incentives can help us question long-held beliefs. For instance, Koriat and Goldsmith (1994) reported that memory accuracy (i.e., the percentage of items that are correctly recalled) is strategically regulated, that is, "subjects can substantially boost their memory accuracy in response to increased accuracy motivation" (p. 307). Koriat and Goldsmith stressed that their results contrast sharply

with the general observation from quantity-oriented research that people cannot improve their memory-quantity performance when given incentives to do so. Participants in a high-accuracy-incentive condition were more accurate than those in a moderate-accuracy-incentive condition ($\eta = .58$, a large effect according to Cohen 1988; calculated from data in Koriat and Goldsmith's 1994 Table 3).

7. Needless to say, the implementation of financial incentives has its own risks. It is, for example, important to ensure that payments are given privately. As a referee correctly pointed out, public payment can be "akin to an announcement of poor test performance and might violate a number of ethical (and, in America, perhaps legal) standards, and is all the more likely to negatively impact mood." Private payment is the standard practice in economics experiments.
8. One reviewer referred us to Frey's (1997) discussion of the hidden costs of extrinsic rewards. Frey's book, as thought provoking and insightful as it often is, takes as its point of departure the same literature that Eisenberger and Cameron (1996) discussed and took issue with. As mentioned, we agree that money does not always work as a motivator, but we believe that more often than not it does. Let us consider Frey's example of professors. Professors who are so engaged in their profession that they teach more than the required hours per week may indeed react with indignation when administrators try to link remuneration more closely to performance and therefore reduce their extra effort. There are, however, also professors who "shirk" (the term used in principal-agent theory) their teaching obligations to do research, consulting, and so forth. In fact, shirking has been identified as the major driver of the inefficiency of educational institutions in the United States (Massy & Zemsky 1994; Ortmann & Squire 2000). While consulting has immediate material payoffs, at most institutions research translates into higher salaries and, possibly more important, payoffs such as the adulation of peers at conferences (Lodge 1995). It is noteworthy that the activities that professors engage in involve by their very nature self-determination, self-esteem, and expression possibility and therefore are particularly susceptible to "crowding out." In contrast, most laboratory tasks do not prominently feature these characteristics.
9. What constitutes deception is not easy to define (see Baumrind 1979; Rosenthal & Rosnow 1991). Economists seem to make the following pragmatic distinction, which we endorse: Telling participants wrong things is deception. Conveying false information to participants, however, is different from not explicitly telling participants the purpose of an experiment, which is not considered deception by either economists (McDaniel & Starmer 1998; Hey 1998) or psychologists known to be opposed to deception (e.g., Baumrind 1985). However, to the extent that absence of full disclosure of the purpose of an experiment violates participants' default assumptions, it can mislead them, and therefore should be avoided.
10. To avoid many "I don't know" responses, we asked economists to estimate how often psychologists in general (rather than researchers in JDM) use various practices. This may explain why their estimates for the use of deception were so high.
11. There are also regulatory standards in psychology – possibly the best examples are the treatment group experiments and null-hypothesis testing (see Danziger 1990). Null-hypothesis testing was, and to a large extent remains, a self-imposed requirement in psychology despite continuous controversy about its use. How is null-hypothesis testing different from the key variables of experimental design considered here? Gigerenzer and Murray (1987) argued that "the inference revolution unified psychology by prescribing a common method, in the absence of a common theoretical perspective" (p. 22). One may speculate that null-hypothesis testing still predominates in psychology because abandoning it may be perceived as abandoning the unification of psychological methodology. The key variables of experimental design considered in this article have never filled this role.

References

Letters "a" and "r" appearing before authors' initials refer to target article and response, respectively.

- Adair, J. G., Dushenko, T. W. & Lindsay, R. C. L. (1985) Ethical regulations and their impact on research practice. *American Psychologist* 40:59-72. [aRH]
- Aitkenhead, M. & Dordoy, J. (1985) What the subjects have to say. *British Journal of Social Psychology* 24:293-305. [aRH]
- Allen, D. F. (1983) Follow-up analysis of use of forewarning and deception in psychological experiments. *Psychological Reports* 52:899-906. [aRH]
- Allison, S. T. & Messick, D. M. (1990) Social decision heuristics in the use of shared resources. *Journal of Behavioral Decision Making* 3:195-204. [aRH]
- American Psychological Association (1992) Ethical principles of psychologists and code of conduct. *American Psychologist* 47:1597-611. [aRH]
- Balzer, W. K., Doherty, M. E. & O'Connor, R. (1989) Effects of cognitive feedback on performance. *Psychological Bulletin* 106:410-33. [aRH]
- Bar-Hillel, M. & Fischhoff, B. (1981) When do base rates affect predictions? *Journal of Personality and Social Psychology* 41:671-80. [aRH]
- Barro, R. J. (1990) *Macro-economic policy*. Harvard University Press. [aRH]
- Baumrind, D. (1964) Some thoughts on ethics of research. After reading Milgram's "Behavioral study of obedience." *American Psychologist* 19:421-23. [aRH]
- (1971) Principles of ethical conduct in the treatment of subjects: Reaction to the draft report of the Committee on Ethical Standards in Psychological Research. *American Psychologist* 26:887-96. [aRH]
- (1979) IRBs and social science research: The costs of deception. *IRB: A Review of Human Subjects Research* 1:1-14. [aRH]
- (1985) Research using intentional deception: Ethical issues revisited. *American Psychologist* 40:165-74. [aRH]
- Beach, L. R. & Phillips, L. D. (1967) Subjective probabilities inferred from estimates and bets. *Journal of Experimental Psychology* 75:354-59. [aRH]
- Beattie, J. & Loomes, G. (1997) The impact of incentives upon risky choice experiments. *Journal of Risk and Uncertainty* 14:155-68. [aRH]
- Beeler, J. D. & Hunton, J. E. (1997) The influence of compensation method and disclosure level on information search strategy and escalation of commitment. *Journal of Behavioral Decision Making* 10:77-91. [aRH]
- Berg, J. E., Dickhaut, J. W. & McCabe, K. A. (1995) Trust, reciprocity, and social history. *Games and Economic Behavior* 10:122-42. [ASGi, aRH]
- Berg, J. E., Dickhaut, J. W. & O'Brien, J. R. (1985) Preference reversal and arbitrage. *Research in Experimental Economics* 3:31-72. [aRH]
- Binmore, K. (1994) *Playing fair*. MIT Press. [GWH, arRH]
- (1999) Why experiment in economics? *The Economic Journal* 109:16-24. [aRH]
- Birnbaum, M. & Mellers, B. A. (1983) Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality and Social Psychology* 45:792-804. [aRH]
- Bonetti, S. (1998) Experimental economics and deception. *Journal of Economic Psychology* 19:377-95. [aRH]
- Bower, B. (1997) Null science: Psychology's statistical status quo draws fire. *Science News* 151:356-57. [aRH]
- Brehmer, B. (1980) In one word: Not from experience. *Acta Psychologica* 45:223-41. [aRH]
- (1992) Dynamic decision making: Human control of complex systems. *Acta Psychologica* 81:211-41. [aRH]

- Brehmer, B. (1996) Man as a stabiliser of systems: From static snapshots of judgment processes to dynamic decision making. *Thinking and Reasoning* 2:225–38. [aRH]
- Bröder, A. (1998) Deception can be acceptable: Comment on Ortmann and Hertwig. *American Psychologist* 58:805–806. [aRH]
- Bruce, D. (1985) The how and why of ecological memory. *Journal of Experimental Psychology: General* 114:78–90. [aRH]
- Butera, F., Mugny, G., Legrenzi, P. & Perez, J. A. (1996) Majority and minority influence, task representation and inductive reasoning. *British Journal of Social Psychology* 35:123–36. [aRH]
- Camerer, C. F. (1990) Do markets correct biases in probability judgment? Evidence from market experiments. In: *Advances in behavioral economics*, ed. L. Green & J. Kagel. Ablex. [aRH]
- (1995) Individual decision making. In: *The handbook of experimental economics*, ed. J. H. Kagel & A. E. Roth. Princeton University Press. [GWH, aRH, DJZ]
- (1997) Rules for experimenting in psychology and economics, and why they differ. In: *Understanding strategic interaction. Essays in honor of Reinhard Selten*, ed. W. Albers, W. Güth, P. Hammerstein, B. Moldovanu & E. van Damme. Springer. [aRH]
- Camerer, C. F. & Ho, T. (1999) Experience-weighted attraction learning (EWA) in “normal-form” games: A unifying approach. *Econometrica* 67:827–74. [aRH, RK]
- Camerer, C. F. & Hogarth, R. M. (1999) The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty* 19:7–42. [aRH, RMH, AK]
- Camerer, C. F., Loewenstein, G. & Weber, M. (1989) The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy* 97:1232–54. [aRH]
- Cameron, J. & Pierce, W. D. (1994) Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational Research* 64:363–423. [aRH]
- Cameron, L. (1999) Raising the stakes in the ultimatum game: Experimental evidence from Indonesia. *Economic Inquiry* 37(1):47–59. [JH]
- Christensen, L. (1977) The negative subject: Myth, reality, or a prior experimental experience effect? *Journal of Personality and Social Psychology* 35:392–400. [aRH]
- (1988) Deception in psychological research: When is its use justified? *Personality and Social Psychology Bulletin* 14:664–75. [aRH]
- Chu, Y. P. & Chu, R. L. (1990) The subsidence of preference reversals in simplified and marketlike experimental settings: A note. *The American Economic Review* 80:902–11. [aRH]
- Cohen, J. (1962) The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology* 69:145–53. [aRH]
- (1988) *Statistical power analysis for the behavioral sciences*, 2nd edition. Erlbaum. [aRH]
- Connolly, T. (1988) Hedge-clipping, tree-felling, and the management of ambiguity. In: *Managing the challenge of ambiguity and change*, ed. M. B. McCaskey, L. R. Pondy & H. Thomas. Wiley. [aRH]
- Cook, T. D., Bean, J. R., Calder, B. J., Frey, R., Krovetz, M. L. & Reisman, S. R. (1970) Demand characteristics and three conceptions of the frequently deceived subject. *Journal of Personality and Social Psychology* 14:185–94. [aRH]
- Cosmides, L. & Tooby, J. (1996) Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition* 58:1–73. [aRH, DSL]
- Creyer, E. H., Bettman, J. R. & Payne, J. W. (1990) The impact of accuracy and effort feedback and goals on adaptive decision behavior. *Journal of Behavioral Decision Making* 3:1–16. [aRH]

- Danziger, K. (1990) *Constructing the subject. Historical origins of psychological research*. Cambridge University Press. [GG, aRH]
- Davies, M. F. (1992) Field dependence and hindsight bias: Cognitive restructuring and the generation of reasons. *Journal of Research in Personality* 26:58-74. [aRH]
- Davis, D. D. & Holt, C. A. (1993) *Experimental economics*. Princeton University Press. [GWH, arRH]
- Dawes, R. M. (1988) *Rational choice in an uncertain world*. Harcourt-Brace Jovanovich. [aRH]
- (1996) The purpose of experiments: Ecological validity versus comparing hypotheses. *Behavioral and Brain Sciences* 19:20. [aRH]
- Deci, E. L., Koestner, R. & Ryan, R. M. (1999a) Meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* 125:627-68. [aRH]
- (1999b) The undermining effect is a reality after all - extrinsic rewards, task interest, and self-determination: Reply to Eisenberger, Pierce, and Cameron (1999) and Lepper, Henderlong, and Gingras (1999). *Psychological Bulletin* 125:692-700. [aRH]
- Dickhaut, J., Hubbard, J. & McCabe, K. (1995) Trust, reciprocity, and interpersonal history: Fool me once, shame on you, fool me twice, shame on me. Working paper, University of Minnesota. [aRH]
- Diehl, E. & Serman, J. D. (1995) Effects of feedback complexity on dynamic decision making. *Organizational Behavior and Human Decision Processes* 62:198-215. [aRH]
- Diener, E. & Crandall, R. (1978) *Ethics in social and behavioral research*. University of Chicago Press. [aRH]
- Duhem, P. (1953) Physical theory and experiment. In: *Readings in the philosophy of science*, ed. H. Feigl & M. Brodbeck. Appleton-Century-Crofts. [aRH]
- Dyer, D. & Kagel, J. H. (1996) Bidding in common value auctions: How the commercial construction industry corrects for the winners' curse. *Management Science* 42:1463-75. [aRH]
- Edwards, W. (1961) Costs and payoffs are instructions. *Psychological Review* 68:275-84. [aRH]
- (1962) Dynamic decision theory and probabilistic information processing. *Human Factors* 4:59-73. [aRH]
- Eisenberger, R. & Cameron, J. (1996) Detrimental effects of reward: Reality or myth? *American Psychologist* 51:1153-66. [aRH]
- Eisenberger, R., Pierce, W. D. & Cameron, J. (1999) Effects of reward on intrinsic motivation - negative, neutral, and positive: Comment on Deci, Koestner, and Ryan (1999). *Psychological Bulletin* 125:677-91. [aRH]
- Epley, N. & Huff, C. (1998) Suspicion, affective response, and educational benefit as a result of deception in psychology research. *Personality and Social Psychology Bulletin* 24:759-68. [aRH]
- Epstein, Y. M., Suedfeld, P. & Silverstein, S. J. (1973) The experimental contract: Subjects' expectations of and reactions to some behaviors of experimenters. *American Psychologist* 28:212-21. [aRH]
- Evans, J. St. B. T. & Bradshaw, H. (1986) Estimating sample-size requirements in research design: A study of intuitive statistical judgment. *Current Psychological Research and Reviews* 5:10-19. [aRH]
- Evans, J. St. B. T., Newstead, S. E. & Byrne, R. M. J. (1993) *Human reasoning: The psychology of deduction*. Erlbaum. [aRH]
- Evans, J. St. B. T., Over, D. E. & Manktelow, K. I. (1993) Reasoning, decision making and rationality. *Cognition* 49:165-87. [aRH]
- Fillenbaum, S. (1966) Prior deception and subsequent experimental performance: The "faithful" subject. *Journal of Personality and Social Psychology* 4:532-37. [aRH]

- Fischhoff, B., Slovic, P. & Lichtenstein, S. (1979) Subjective sensitivity analysis. *Organizational Behavior and Human Performance* 23:339–59. [aRH]
- Fisher, C. B. & Fyrberg, D. (1994) Participant partners: College students weigh the costs and benefits of deceptive research. *American Psychologist* 49:417–27. [aRH]
- Frey, B. S. (1997) *Not just for the money: An economic theory of personal motivation*. Edward Elgar. [aRH]
- Garnham, A. & Oakhill, J. (1994) *Thinking and reasoning*. Blackwell. [aRH]
- Geller, D. M. (1978) Involvement in role-playing simulations: A demonstration with studies on obedience. *Journal of Personality and Social Psychology* 36:219–35. [aRH]
- Gigerenzer, G. (1996) On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review* 103:592–96. [aRH]
- Gigerenzer, G. & Hoffrage, U. (1995) How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review* 102:684–704. [GG, aRH]
- Gigerenzer, G. & Hug, K. (1992) Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition* 43:127–71. [GG, aRH]
- Gigerenzer, G. & Murray, D. J. (1987) *Cognition as intuitive statistics*. Erlbaum. [aRH]
- Glass, G. V., McGaw, B. & Smith, M. L. (1981) *Meta-analysis in social research*. Sage. [aRH]
- Goeree, J. K. & Holt, C. A. (in press) Ten little treasures of game theory and ten intuitive contradictions. *American Economic Review*. [aRH]
- Goldstein, W. M. & Weber, E. U. (1997) Content and discontent: Indications and implications of domain specificity in preferential decision making. In: *Research on judgment and decision making: Currents, connections, and controversies*, ed. W. M. Goldstein & R. M. Hogarth. Cambridge University Press. [aRH]
- Greenwood, J. D. (1983) Role-playing as an experimental strategy in social psychology. *European Journal of Social Psychology* 13:235–54. [aRH]
- Grether, D. M. (1980) Bayes rule as a description model: The representativeness heuristic. *Quarterly Journal of Economics* 95:537–57. [aRH]
- (1992) Testing Bayes rule and the representativeness heuristic: Some experimental evidence. *Journal of Economic Behavior and Organization* 17:31–57. [aRH]
- Grether, D. M. & Plott, C. R. (1979) Economic theory of choice and the preference reversal phenomenon. *American Economic Review* 69:623–38. [FG, aRH]
- Griggs, R. A. & Cox, J. R. (1982) The elusive thematic materials effect in the Wason's selection task. *British Journal of Psychology* 73:407–20. [aRH]
- Hammond, K. R., Stewart, T. R., Brehmer, B. & Steinman, D. O. (1975) Social judgment theory. In: *Human judgment and decision processes*, ed. M. F. Kaplan & S. Schwartz. Academic Press. [aRH]
- Harless, D. W. & Camerer, C. F. (1994) The predictive utility of generalized expected utility theories. *Econometrica* 62:1251–89. [aRH]
- Harrison, G. W. (1989) Theory and misbehavior of first-price auctions. *American Economic Review* 79:749–62. [GWH, aRH]
- (1992) Theory and misbehavior of first-price auctions: Reply. *American Economic Review* 82:1426–43. [GWH, aRH]
- (1994) Expected utility theory and the experimentalists. *Empirical Economics* 19:223–53. [aRH]
- (1999) *Experimental economics and contingent valuation*. Working Paper 96–10, Division of Research, The Darla Moore School of Business, University of South Carolina (<http://theweb.badm.sc.edu/glenn/eecv.pdf>). [aRH]
- Harrison, G. W. & Rutstroem, E. E. (in press) Experimental evidence on the existence of hypothetical bias in value elicitation methods. In: *Handbook of experimental economics results*, ed. C. R. Plott & V. L. Smith. Elsevier. [aRH]
- Hedges, L. V. & Olkin, I. (1985) *Statistical methods for meta-analysis*. Academic Press. [aRH]

- Hell, W., Gigerenzer, G., Gauggel, S., Mall, M. & Müller, M. (1988) Hindsight bias: An interaction of automatic and motivational factors? *Memory and Cognition* 16:533-38. [aRH]
- Hertwig, R. & Gigerenzer, G. (1999) The "conjunction fallacy" revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making* 12:275-305. [GG, arRH]
- Hey, J. D. (1991) *Experiments in economics*. Blackwell. [aRH]
- (1998) Experimental economics and deception: A comment. *Journal of Economic Psychology* 19:397-401. [aRH]
- Hilton, D. (1995) The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin* 118:248-71. [aRH]
- Hinsz, V. B., Tindale, R. S., Nagao, D. H., Davis, J. H. & Robertson, B. A. (1988) The influence of the accuracy of individuating information on the use of base rate information in probability judgment. *Journal of Experimental Social Psychology* 24:127-45. [aRH]
- Hoffman, E., McCabe, K. A. & Smith, V. L. (1996) Social distance and other-regarding behavior in dictator games. *American Economic Review* 86:653-60. [aRH]
- Hoffrage, U. & Hertwig, R. (1999) Hindsight bias: A price worth paying for fast and frugal memory. In: *Simple heuristics that make us smart*, ed. G. Gigerenzer, P. M. Todd & the ABC Research Group. Oxford University Press. [aRH]
- Hogarth, R. M. (1981) Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin* 90:197-217. [aRH, RMH]
- Hogarth, R. M., Gibbs, B. J., McKenzie, C. R. M. & Marquis, M. A. (1991) Learning from feedback: Exactness and incentives. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17:734-52. [aRH]
- Hogarth, R. M. & Reder, M. W. (1987) Introduction: Perspectives from economics and psychology. In: *Rational choice: The contrast between economics and psychology*, ed. R. M. Hogarth & M. W. Reder. University of Chicago Press. [aRH]
- Holland, J. S. & Kleinmuntz, D. N. (1994) Factors influencing the use of internal summary evaluations versus external information in choice. *Journal of Behavioral Decision Making* 7:79-102. [aRH]
- Hunter, J. E. & Schmidt, F. L. (1990) *Methods of meta-analysis: Correcting error and bias in research findings*. Sage. [aRH]
- Irwin, J. R., McClelland, G. H. & Schulze, W. D. (1992) Hypothetical and real consequences in experimental auctions for insurance against low-probability risks. *Journal of Behavioral Decision Making* 5:107-16. [aRH]
- Jamal, K. & Sunder, S. (1991) Money vs. gaming: Effects of salient monetary payments in double oral auctions. *Organizational Behavior and Human Decision Processes* 49: 151-66. [aRH]
- Jenkins, G. D., Jr., Mitra, A., Gupta, N. & Shaw, J. D. (1998) Are financial incentives related to performance? A meta-analytic review of empirical research. *Journal of Applied Psychology* 83:777-87. [aRH]
- Kahneman, D., Slovic, P. & Tversky, A. (1982) *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press. [aRH, RMH]
- Kahneman, D. & Tversky, A. (1973) On the psychology of prediction. *Psychological Review* 80:237-51. [arRH, CRMM]
- (1996) On the reality of cognitive illusions: A reply to Gigerenzer's critique. *Psychological Review* 103:582-91. [aRH]
- Kelman, H. C. (1967) Human use of human subjects: The problem of deception in social psychological experiments. *Psychological Bulletin* 67:1-11. [aRH]
- Kimmel, A. J. (1996) *Ethical issues in behavioral research*. Blackwell. [aRH]
- (1998) In defense of deception. *American Psychologist* 53:803-805. [aRH]
- Klayman, J. (1988) On the how and why (not) of learning from outcomes. In: *Human judgment: The SJT view*, ed. B. Brehmer & C. R. B. Joyce. North-Holland. [aRH]

- Koehler, J. J. (1996) The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences* 19:1–53. [aRH]
- Koriat, A. & Goldsmith, M. (1994) Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General* 123:297–315. [aRH]
- (1996) Memory metaphors and the real-life/laboratory controversy: Correspondence versus storehouse conceptions of memory. *Behavioral and Brain Sciences* 19:167–228. [aRH]
- Kreps, D. M. (1990) *A course in microeconomic theory*. Princeton University Press. [aRH]
- Krupat, E. (1977) A re-assessment of role playing as a technique in social psychology. *Personality and Social Psychology Bulletin* 3:498–504. [aRH]
- Krupat, E. & Garonzik, R. (1994) Subjects' expectations and the search for alternatives to deception in social psychology. *British Journal of Social Psychology* 33:211–22. [aRH]
- Ledyard, J. O. (1995) Public goods: A survey of experimental research. In: *The handbook of experimental economics*, ed. J. H. Kagel & A. E. Roth. Princeton University Press. [aRH]
- Lepper, M. R., Greene, D. & Nisbett, R. E. (1973) Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology* 28:129–37. [aRH]
- Lepper, M. R., Henderlong, J. & Gingras, I. (1999) Understanding the effects of extrinsic rewards on intrinsic motivation – uses and abuses of meta-analysis: Comment on Deci, Koestner, and Ryan (1999). *Psychological Bulletin* 125:669–76. [aRH]
- Lepper, M. R., Keavney, M. & Drake, M. (1996) Intrinsic motivation and extrinsic rewards: A commentary on Cameron and Pierce's meta-analysis. *Review of Educational Research* 66:5–32. [aRH]
- Levin, I. P., Chapman, D. P. & Johnson, R. D. (1988) Confidence in judgments based on incomplete information: An investigation using both hypothetical and real gambles. *Journal of Behavioral Decision Making* 1:29–41. [aRH]
- Lichtenstein, S. & Slovic, P. (1971) Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology* 89:46–55. [aRH]
- Lindeman, S. T., van den Brink, W. P. & Hoogstraten, J. (1988) Effect of feedback on base-rate utilization. *Perceptual and Motor Skills* 67:343–50. [aRH]
- Lipsey, R. (1979) *An introduction to positive economics*, 5th edition. Weidenfeld and Nicholson. [aRH]
- Lodge, D. (1995) *Small world*. Penguin Books. [aRH]
- Loewenstein, G. (1999) Experimental economics from the vantage-point of behavioural economics. *The Economics Journal* 109:F25–F34. [aRH, AK, DJZ]
- Loomes, G. (1999) Some lessons from past experiments and some challenges for the future. *The Economic Journal* 109:35–45. [aRH]
- Lopes, L. L. (1987) Procedural debiasing. *Acta Psychologica* 64:167–85. [aRH]
- (1994) Psychology and economics: Perspectives on risk, cooperation, and the marketplace. *Annual Review of Psychology* 45:197–227. [aRH]
- MacCoun, R. J. & Kerr, N. L. (1987) Suspicion in the psychological laboratory: Kelman's prophecy revisited. *American Psychologist* 42:199. [aRH]
- Manis, M., Dovalina, I., Avis, N. E. & Cardoze, S. (1980) Base rates can affect individual predictions. *Journal of Personality and Social Psychology* 38:231–48. [aRH]
- Massy, W. F. & Zemsky, R. (1994) Faculty discretionary time: Departments and the "academic ratchet." *Journal of Higher Education* 65:1–22. [aRH]
- McDaniel, T. & Starmer, C. (1998) Experimental economics and deception: A comment. *Journal of Economic Psychology* 19:403–409. [aRH]
- Medin, D. L. & Bettger, J. G. (1991) Sensitivity to changes in base-rate information. *American Journal of Psychology* 104:311–32. [aRH]

- Medin, D. L. & Edelson, S. M. (1988) Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General* 117:68–85. [aRH]
- Meehl, P. E. (1978) Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology* 46:806–34. [CFB, aRH]
- Mellers, B. A., Berretty, P. M. & Birnbaum, M. H. (1995) Dominance violations in judged prices of two- and three-outcome gambles. *Journal of Behavioral Decision Making* 8:201–16. [aRH]
- Mellers, B. A., Hertwig, R. & Kahneman, D. (2001) Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science* 12:269–75. [aRH]
- Mixon, D. (1972) Instead of deception. *Journal for the Theory of Social Behaviour* 2:145–77. [aRH]
- Nelson, T. E., Biernat, M. R. & Manis, M. (1990) Everyday base rates (sex stereotypes): Potent and resilient. *Journal of Personality and Social Psychology* 59:664–75. [aRH]
- Newberry, B. H. (1973) Truth telling in subjects with information about experiments: Who is being deceived? *Journal of Personality and Social Psychology* 25:369–74. [aRH]
- Nicks, S. D., Korn, J. H. & Mainieri, T. (1997) The rise and fall of deception in social psychology and personality research, 1921 to 1994. *Ethics and Behavior* 7:69–77. [aRH]
- Oliansky, A. (1991) A confederate's perspective on deception. *Ethics and Behavior* 1:253–58. [aRH]
- Olson, M. J. & Budescu, D. V. (1997) Patterns of preference for numerical and verbal probabilities. *Journal of Behavioral Decision Making* 10:117–31. [aRH]
- Ordóñez, L. D., Mellers, B. A., Chang, S. J. & Roberts, J. (1995) Are preference reversals reduced when made explicit? *Journal of Behavioral Decision Making* 8:265–77. [aRH]
- Orne, M. T. (1962) On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist* 17:776–83. [aRH]
- Ortmann, A. & Colander, D. C. (1997) A simple principal-agent experiment for the classroom. *Economic Inquiry* 35:443–50. [aRH]
- Ortmann, A. & Gigerenzer, G. (1997) Reasoning in economics and psychology: Why social context matters. *Journal of Institutional and Theoretical Economics* 153(4):700–10. [aRH]
- Ortmann, A. & Hertwig, R. (1997) Is deception acceptable? *American Psychologist* 52:746–47. [aRH]
- (1998) The question remains: Is deception acceptable? *American Psychologist* 53:806–807. [aRH]
- Ortmann, A. & Squire, R. C. (2000) A game-theoretic explanation of the administrative lattice in institutions of higher learning. *Journal of Economic Behavior and Organization* 43:377–91. [aRH]
- Ortmann, A. & Tichy, L. (1999) Understanding gender effects in the laboratory: Evidence from Prisoner's Dilemma games. *Journal of Economic Behavior and Organization* 39:327–39. [aRH]
- Osborne, M. J. & Rubinstein, A. (1990) *Bargaining and markets*. Academic Press. [aRH]
- Payne, J. W., Bettman, J. R. & Johnson, E. J. (1992) Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology* 43:87–131. [aRH]
- Payne, J. W., Bettman, J. R. & Luce, M. F. (1996) When time is money: Decision behavior under opportunity-cost time pressure. *Organizational Behavior and Human Decision Processes* 66:131–52. [aRH]
- Pelham, B. W. & Neter, E. (1995) The effect of motivation of judgment depends on the difficulty of the judgment. *Journal of Personality and Social Psychology* 68:581–94. [aRH]

- Prendergast, C. (1999) The provision of incentives in firms. *Journal of Economic Literature* 37:7-63. [aRH]
- Quine, W. V. (1953a) *From a logical point of view*. Harvard University Press. [aRH]
- (1953b) Two dogmas of empiricism. In: *From a logical point of view*. Harvard University Press. [ASGi]
- Rabin, M. (1998) Psychology and economics. *Journal of Economic Literature* 36:11-46. [aRH]
- Rapoport, A. & Wallsten, T. S. (1972) Individual decision behavior. *Annual Review of Psychology* 23:131-76. [aRH]
- Riecken, H. W. (1962) A program for research on experiments in social psychology. In: *Decisions, values and groups, vol. 2*, ed. N. F. Washburne. Pergamon. [aRH]
- Ring, K. (1967) Experimental social psychology: Some sober questions about some frivolous values. *Journal of Experimental Social Psychology* 3:113-23. [aRH]
- Rosenthal, R. (1990) How are we doing in soft psychology? *American Psychologist* 45:775-77. [FJG-W, aRH]
- Rosenthal, R. & Rosnow, R. L. (1991) *Essentials of behavioral research: Methods and data analysis, 2nd edition*. McGraw Hill. [aRH]
- Rossi, J. S. (1990) Statistical power of psychological research: What have we gained in 20 years? *Journal of Consulting and Clinical Psychology* 58:646-56. [aRH]
- Roth, A. E. (1995) Introduction to experimental economics. In: *The handbook of experimental economics*, ed. J. H. Kagel & A. E. Roth. Princeton University Press. [ASGo, aRH, DSL]
- Roth, A. E. & Erev, I. (1995) Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior* 8:164-212. [aRH]
- Rubin, Z. (1985) Deceiving ourselves about deception: Comment on Smith and Richardson's "amelioration of deception and harm in psychological research." *Journal of Personality and Social Psychology* 48:252-53. [aRH]
- Russell, R. R. & Wilkinson, M. (1979) *Microeconomics: A synthesis of modern and neoclassical theory*. Wiley. [aRH]
- Schmidt, F. L. (1992) What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist* 47:1173-81. [aRH]
- Schotter, A., Weiss, A. & Zapater, I. (1996) Fairness and survival in ultimatum and dictatorship games. *Journal of Economic Behavior and Organization* 31:37-56. [aRH]
- Schulz, D. P. (1969) The human subject in psychological research. *Psychological Bulletin* 72:214-28. [aRH]
- Schwartz, B. (1982) Reinforcement-induced behavioral stereotypy: How not to teach people to discover rules. *Journal of Experimental Psychology: General* 111:23-59. [aRH]
- (1989) *Psychology of learning and behavior, 3rd edition*. Norton. [RK]
- Sedlmeier, P. & Gigerenzer, G. (1989) Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin* 105:309-16. [aRH]
- Selten, R. (1998) Aspiration adaptation theory. *Journal of Mathematical Psychology* 42:191-214. [aRH]
- Sharpe, D., Adair, J. G. & Roese, N. J. (1992) Twenty years of deception research: A decline in subjects' trust? *Personality and Social Psychology Bulletin* 18:585-90. [aRH]
- Sieber, J. E., Iannuzzo, R. & Rodriguez, B. (1995) Deception methods in psychology: Have they changed in 23 years? *Ethics and Behavior* 5:67-85. [aRH]
- Sieber, J. E. & Saks, M. J. (1989) A census of subject pool characteristics and policies. *American Psychologist* 44:1053-61. [aRH]
- Simon, H. A. (1957) *Models of man*. Wiley. [aRH]
- Smith, A. (1759/1976) *The theory of moral sentiments*. Oxford University Press. [rRH]
- Smith, M. J. (1982) *Evolution and the theory of games*. Cambridge University Press. [rRH]
- Smith, S. M. & Levin, I. P. (1996) Need for cognition and choice framing effects. *Journal of Behavioral Decision Making* 9:283-90. [aRH]

- Smith, S. S. & Richardson, D. (1983) Amelioration of deception and harm in psychological research: The important role of debriefing. *Journal of Personality and Social Psychology* 44:1075-82. [aRH]
- Smith, V. L. (1991) Rational choice: The contrast between economics and psychology. *Journal of Political Economy* 99:877-97. [aRH]
- Smith, V. L. & Walker, J. M. (1993a) Monetary rewards and decision cost in experimental economics. *Economic Inquiry* 31:245-61. [aRH]
- (1993b) Rewards, experience and decision costs in first price auctions. *Economic Inquiry* 31:237-45. [aRH]
- Stang, D. J. (1976) Ineffective deception in conformity research: Some causes and consequences. *European Journal of Social Psychology* 6:353-67. [aRH]
- Starmer, C. (1999) Experimental economics: Hard science or wasteful tinkering? *The Economic Journal* 109:5-15. [aRH]
- Stone, D. N. & Schkade, D. A. (1994) Effects of attribute scales on process and performance in multiattribute choice. *Organizational Behavior and Human Decision Processes* 59:261-87. [aRH]
- Stone, D. N. & Ziebart, D. A. (1995) A model of financier incentive effects in decision making. *Organizational Behavior and Human Decision Processes* 61:250-61. [aRH]
- Taylor, K. M. & Shepperd, J. A. (1996) Probing suspicion among participants in deception research. *American Psychologist* 51:886-87. [aRH]
- Thaler, R. (1987) The psychology of choice and the assumptions of economics. In: *Laboratory experimentation in economics: Six points of view*, ed. A. E. Roth. Cambridge University Press. [aRH, DSL]
- Tirole, J. (1988) *The theory of industrial organization*. MIT Press. [aRH]
- Tversky, A. & Kahneman, D. (1981) The framing of decisions and the rationality of choice. *Science* 211:453-58. [aRH, DSL]
- (1983) Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review* 90:293-315. [aRH]
- (1987) Rational choice and the framing of decisions. In: *Rational choice: The contrast between economics and psychology*, ed. R. M. Hogarth & M. W. Reder. University of Chicago Press. [aRH]
- Van Wallendael, L. R. (1995) Implicit diagnosticity in an information-buying task: How do we use the information that we bring with us to a problem? *Journal of Behavioral Decision Making* 8:245-64. [aRH]
- Van Wallendael, L. R. & Guignard, Y. (1992) Diagnosticity, confidence, and the need for information. *Journal of Behavioral Decision Making* 5:25-37. [aRH]
- Von Winterfeldt, D. & Edwards, W. (1982) Cost and payoffs in perceptual research. *Psychological Bulletin* 91:609-22. [aRH]
- Wallsten, T. S. (1972) Conjoint-measurement framework for the study of probabilistic information processing. *Psychological Review* 79:245-60. [aRH]
- (1976) Using conjoint-measurement models to investigate a theory about probabilistic information processing. *Journal of Mathematical Psychology* 14:144-85. [aRH]
- (1982) Research through deception (letter to the editor). *New York Times Magazine*, October 1982. [aRH]
- Wallsten, T. S. & Barton, C. (1982) Processing probabilistic multidimensional information for decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 8:361-84. [aRH]
- Weber, S. J. & Cook, T. D. (1972) Subject effects in laboratory research: An examination of subjects role, demand characteristics, and valid inference. *Psychological Bulletin* 77:273-95. [aRH]
- Wilcox, N. T. (1993) Lottery choice: Incentives, complexity and decision time. *Economic Journal* 103:1397-417. [aRH]

- Winkler, R. L. & Murphy, A. H. (1973) Experiments in the laboratory and the real world. *Organizational Behavior and Human Performance* 10:252-70. [aRH]
- Yaniv, I. & Schul, Y. (1997) Elimination and inclusion procedures in judgment. *Journal of Behavioral Decision Making* 10:211-20. [aRH]
- Zwick, R., Erev, I. & Budescu, D. V. (1999) The psychological and economical perspectives on human decisions in social and interactive contexts. In: *Games and human behavior: Essays in honor of Amnon Rapoport*, ed. D. V. Budescu, I. Erev & R. Zwick. Erlbaum. [aRH]