

---

# Economists' and Psychologists' Experimental Practices: How They Differ, Why They Differ, and How They Could Converge

RALPH HERTWIG AND ANDREAS ORTMANN

## 1. INTRODUCTION

We all value the notion of interdisciplinarity. But, principles notwithstanding, dialogues across borders do not occur very often. Why is this so? There are myriad hurdles that stand in the way of transcending the limits of our disciplines, ranging from the incommensurability of theoretical frameworks and languages to differences in subject matters to mundane institutional barriers (e.g., interdisciplinary work is not a good bet for getting tenured). Yet another hindrance is the fact that different disciplines often employ different tools in the production of scientific knowledge. And, unfortunately, we often move quickly from recognizing the other tribe's different methodological tools and conventions to a suspicion of their epistemological soundness. Needless to say this suspicion tends to grow larger when the findings generated by alien practices challenge one's own long-held beliefs and assumptions. Interactions between experimental economists and psychologists are not exempt from this phenomenon. Fortunately, however, one's own beliefs about proper methodological conduct, as much as suspicions about the other tribe's practices, can be empirically scrutinized.

The goal of this chapter is to subject some of our beliefs to such an analysis. Specifically, we will discuss economists' and psychologists' different conceptions of proper experimentation in terms of two key variables of experimental design—the use (or lack thereof) of financial incentives and deception. Both variables have been suggested to epitomize the conflicting methodological views of economists and psychologists. Moreover, these variables figure prominently in economists'

We would like to thank Juan Carrillo and Daniel Gilbert for constructive comments. Special thanks are due to Anita Todd for improving the readability of the manuscript, and the Deutsche Forschungsgemeinschaft for financial support to the first author (Forschungsstipendium He 2768/6-1).

debates about the reliability of empirical results from psychology (e.g., Grether and Plott, 1979). We will address two questions: First, do psychologists and economists, in fact, realize these variables differently, and if so, why? Second, do the different realizations matter in terms of the outcome of the experiments? We conclude the discussion of each of these questions with a policy recommendation.

Two caveats are in order. The fact that we focus on two key variables of experimental design does not make others irrelevant. In fact, elsewhere we have discussed additional variables (Hertwig and Ortmann, 2001). Moreover, 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 do not apply, or apply to a much lesser degree, to the research practices in other fields of psychology (such as biological psychology or psychophysics).

## 2. WHY FINANCIAL INCENTIVES AND WHY DECEPTION?

The few exchanges over methodological issues between economists and psychologists have been intimately linked to a contentious debate. This debate concerns the question of to what extent, if at all, individual decision-makers depart from the principles of rational economic behavior. Psychologists have accumulated experimental evidence since the 1970s that suggests that 'behavioral assumptions employed by economists are simply wrong' (Grether, 1978, p. 70; see Camerer, 1995; Hogarth and Reder, 1987; and Rabin, 1998, for a summary). Economists have countered by challenging the relevance of this evidence. One prominent reply is that even if individuals' decision-making is flawed, the market will correct for errors and biases (e.g., by averaging out violations that may essentially be random mistakes, or by driving agents who make mistakes from the market; see Camerer, 1990, for a summary and discussion of these and other arguments).

Another prominent reply, and the one that is most relevant for the present discussion, questions the validity of the experimental evidence itself. Experimentally observed violations of rational choice models, so the argument goes, could be peculiar to the methodological customs and rituals of psychologists. David Grether and Charles Plott (1979) are among the economists who have given voice to the misgivings about psychologists' methodologies. In a classic article published in the *American Economic Review*, they explored the robustness of one of the violations psychologists had documented, the *preference reversal* phenomenon. Among others, two features of psychological experimentation raised their suspicion: the lack of financial incentives and the use of deception. Here is why:

1. *Mis-specified Incentives*. Almost all economic theory is applied to situations where the agent choosing is seriously concerned or is at least choosing from among options that in some sense matter. . . . Thus, the results of experiments

where subjects may be bored, playing games, or otherwise not motivated, present no immediate challenges to theory.

2. *The Experimenters were Psychologists.* In a very real sense this can be a problem. Subjects nearly always speculate about the purpose of experiments and psychologists have the reputation for deceiving subjects. . . . In order to give the results additional credibility, we felt that the experimental setting should be removed from psychology (Grether and Plott, 1979, pp. 624, 629).

On the basis of their own experimental studies, Grether and Plott (1979) observed that preference reversals survive under experimental conditions that conform to economists' methodological preferences (but see Chu and Chu, 1990). Nevertheless, the lack of performance-based monetary payments (henceforth, *financial incentives*) and the use of deception in psychological experimentation have remained a concern. More than ten years later, the economist Vernon Smith, for instance, concluded that the fact that financial incentives 'are commonly absent in the research of psychologists . . . has made their work vulnerable to the criticism that the results are not meaningful' (Smith 1991, p. 887). In what follows, we examine the claim that incentives are not commonly used in psychological experimentation and then address the question of to what extent financial incentives matter for the results obtained.

### 3. FINANCIAL INCENTIVES: HOW DIVERGENT ARE ECONOMISTS' AND PSYCHOLOGISTS' PRACTICES, AND WHY?

Unlike psychologists, experimental economists who do not use financial incentives are pretty much assured that their results will not be published in respected journals. According to Camerer and Hogarth (1999), during 1970–97, every published experimental study in the *American Economic Review* paid its participants according to performance. What motivates this strict norm? Arguably, the most important reason is the one stated by Grether and Plott (1979). Economic theory applies to situations in which there is something at stake for the decision maker (see also Smith, 1991, p. 887). If nothing or too little is at stake in an experimental setting, so the argument goes, participants may not bother to think carefully about the problem and will respond in an offhand, unreliable fashion.

The rationale behind this belief is that economists think of 'cognitive effort' as a scarce resource that people allocate strategically. If participants are not paid contingent on their performance, then they will not invest the cognitive effort necessary to avoid judgment errors. In contrast, if payoffs are provided that satisfy certain requirements such as 'payoff dominance' (Smith, 1976, 1982), participants will invest cognitive effort.<sup>1</sup> As a consequence, performance variability will

<sup>1</sup> Smith (1982) proposed several precepts to define what is meant by a controlled microeconomic experiment. Harrison (1989, 1992) argued that experiments in economics that provide financial incentives nevertheless often fail to operationalize satisfactorily one of these precepts, namely, 'payoff

be reduced (Davis and Holt, 1993, p. 25), and performance will be closer to the predictions of the economic theory (Smith and Walker, 1993). Other arguments for financial incentives are derivatives of this general argument: for instance, the assumption that the saliency of financial incentives is easier to gauge and implement than most alternative incentives, or the assumption that most of us want more money, and thus there is no satiation over the course of an experiment.

Compared to economists' strict norm, what is the practice in psychology? Though they are frequently a bone of contention, we know very little about how often (performance-based) financial incentives have been used in psychological experimentation. In a first attempt to quantify the actual practice, we examined all the articles published in the *Journal of Behavioral Decision Making (JBDM)* in the years 1988–97. The *JBDM* is a major publication for behavioral decision researchers in psychology. We included 186 studies in the analysis (see Hertwig and Ortmann, 2001, for details of the procedure).

In this large pool of studies, were financial incentives in fact 'commonly absent' (Smith, 1991, p. 887)? Of those 186 studies, about a quarter (48, or 26 percent) employed financial incentives. However, one quarter is likely to be an upper-bound estimate for the prevalence of financial incentives in psychology, because *JBDM* publishes articles at the intersection of psychology, management science, and economics, and experimental economists are on the editorial board. Indeed, in another large sample of 106 empirical studies on Bayesian reasoning, published in a wide cross-section of psychology journals, we (Hertwig and Ortmann, 2001) found that fewer than three percent provided financial incentives (the sample of studies was drawn from a recent review of Bayesian reasoning studies by Koehler, 1996). In sum, while Smith's assertion that financial incentives are commonly absent in psychology sounds harsh, the evidence seems to support him.

Why are financial incentives rarely used by psychologists? The considerations frequently raised in conversations and writing are pragmatic, theoretical, and ethical. On a pragmatic level, some psychologists fear that the cost of experiments will become prohibitive, limiting the ability of young investigators to conduct empirical studies. Regular financial incentives appear even more unacceptable when coupled with the belief that 'our subjects are the usual middle-class achievement-oriented people who wish to provide (maximal performance)' (Dawes 1996, p. 20). Such a belief suggests that financial incentives are redundant.

dominance'. Lack of payoff dominance describes essentially flat maxima, which make it relatively inexpensive for participants not to choose the theoretically optimal action. Thus, for payoffs to satisfy the dominance requirement (Smith, 1976), the difference between the payoff for participants' actual behavior and that for optimal behavior in an experiment needs to be monetarily significant to participants. For instance, if the difference between the payoff for the participant's actual behavior and that for optimal behavior is small (see Harrison, 1994, for examples), one could argue that the payoff decrement participants accept by not behaving optimally is too trivial to be considered meaningful.

A theoretical consideration is that the imposition of a monetary structure on experiments unduly restricts people's wide range of motivations to just one, and may even crowd out intrinsic motivation (see Deci et al., 1999; but also Eisenberger and Cameron, 1996). In addition, incentives in the experimental situation may provide cues to the experimenter's intent ('demand characteristics') and thus bias participants' responses. As a consequence, participants' behavior may conform to the experimenter's hypotheses, for reasons independent of the hypotheses the experimenter intends to test (e.g., Dawes, 1999). Finally, it has been argued that performance-contingent financial incentives have ethical implications. Public payment, for instance, can be akin to an announcement of poor test performance and, thus, might violate a number of ethical standards. We will return to these objections later, but first will consider the question of the effects of financial incentives.

#### 4. FINANCIAL INCENTIVES: DO THE DIFFERENT PRACTICES MATTER?

The impact (or lack thereof) of financial incentives is an important but also contentious issue, with each side of the debate pointing to empirical evidence that supports its particular claims. Tversky and Kahneman (1987, p. 90), for instance, argued that 'experimental findings provide little support' for the view that 'observed failures of rational models are attributable to the cost of thinking and will thus be eliminated by proper incentives . . . . In particular, elementary blunders of probabilistic reasoning . . . are hardly reduced by incentives.' Referring to this conclusion, Smith (1991, p. 887) responded that 'strangely missing, since it is said that there is a "little" evidence, are any of the many citations that could have been offered showing that monetary rewards matter.'

Not surprisingly, the fact that the results are mixed contributes to the suspicion on each side that the other is selectively citing evidence. A lack of review studies that rely on the representative samples of experiments certainly does not allay this skepticism. Our ten-year sample of empirical studies published in *JBDM*, however, should be unbiased insofar as its purpose was to quantify the frequency of financial incentives and not to demonstrate that they either matter or do not matter. In the *JBDM* sample, 48 of 186 studies employed financial incentives, but only 10 of those 48 studies systematically explored the effect of financial incentives (by comparing either a payment to a nonpayment condition or different payment schemes).

In a relatively small sample that nevertheless covers a wide range of research topics, such as framing effects, auctions, evaluations of gambles, preference reversal, and information search, what kind of effects did we find? In the majority of cases in which financial incentives made a difference, they *improved* participants' performance, that is, it brought decisions closer to the predictions of the normative models (see Hertwig and Ortmann's, 2001, Table 2 and their text for the detailed results). In only two cases did financial incentives seem to impair

performance (but one of the two cases was compromised by methodological problems), and in a few cases, they did not affect performance at all.

In other words, what we found is that although financial incentives certainly do not guarantee optimal decisions, in many cases they bring decisions closer to the predictions of the normative models. Equally important, some studies in our sample reported that financial incentives substantially reduce data variability. These results are in line with the central conclusions from a recent survey by Smith and Walker (1993). Their survey, however, did not specify the inclusion criteria for the studies sampled.

Aside from the Smith and Walker study, only a few other recent review articles have explored the effects of financial incentives (e.g., Jenkins et al., 1998; Prendergast, 1999). We focus here on the analysis by Camerer and Hogarth (1999) because they considered studies both from psychology and economics. They compiled a total of 59 studies, using what they call an 'opportunistic sampling' approach (i.e., a nonrandom sampling insofar as they included studies they knew of and that came to their attention). The studies came from three different research domains, namely, 'judgments and decisions', 'games and markets', and 'individual choice'. Taken together, the three research domains yielded mixed results: In 45 percent of the studies, financial incentives did not make a difference; in 40 percent they helped; and in 15 percent their effects were negative. But when we consider the first two domains separately, interesting differences emerge.

As Figure 13.1 shows, the largest effect of financial incentives occurred for 'judgment and decision' studies: In 15 of 28 studies (53 percent), the financial

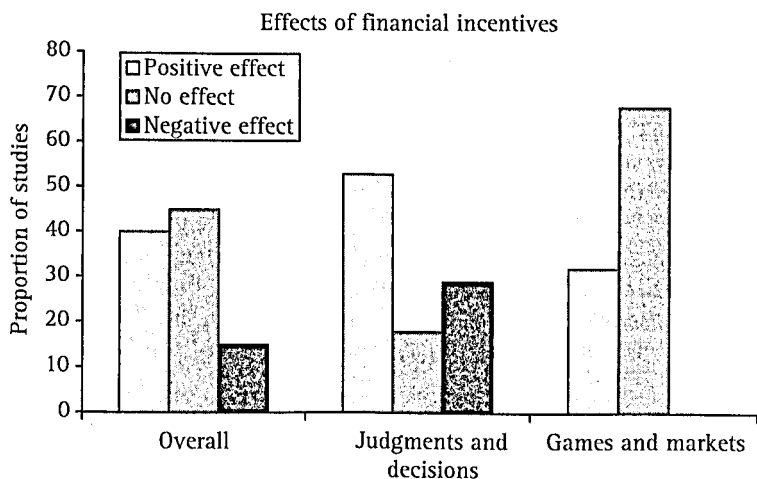


Figure 13.1. The proportion of studies exhibiting various incentive effects in Camerer and Hogarth's (1999) sample of studies. The graphs show the results averaged across all studies (left), and for two subset of studies ('judgments and decisions' vs. 'games and markets' studies)

incentives had positive effects, in 5 (18 percent) they had no effect, and in 8 (29 percent) they had negative effects. Regarding the latter studies, however, Camerer and Hogarth (1999, pp. 6 and 8; emphasis is theirs) concluded that the 'effects are often unclear for various methodological reasons.' Moreover, they reported that in 'many of the studies where incentives did not affect mean performance, added incentives *did* reduce variation.'

In contrast to the 'judgment and decision' studies, Camerer and Hogarth (1999) observed that among the 'game and market' studies the effect of financial incentives appears to be substantially weaker. In only 7 of 22 studies (32 percent) did incentives have positive effects, while in 15 (68 percent) they had no effect (this lack of effect was even more pronounced for the few 'individual choice' studies, see Camerer and Hogarth, 1999, table 2). To the extent that one accepts Camerer and Hogarth's nonrandom sample as representative of the population of judgments, decisions, games, and markets experiments, the results are not without some irony. On the home turf of economists—'game and market' studies—incentives may matter less than it is traditionally assumed. In contrast, on the home turf of psychologists—'judgment and decision' studies—incentives may matter more. While the former conclusion (about the game and market studies) conflicts with Smith and Walker's (1993), the latter (about the judgment and decision studies) is in line with the results observed in our nonopportunistic sample of studies drawn from the *Journal of Behavioral Decision Making* (Hertwig and Ortmann, 2001).

## 5. FINANCIAL INCENTIVES: WHERE TO GO FROM HERE?

The picture that emerges from these results can be summarized as follows: First, financial incentives matter more in some areas than in others (Camerer and Hogarth, 1999). Second, in the area of behavioral decision-making they matter more often than not (see Camerer and Hogarth, 1999; Hertwig and Ortmann, 2001), while the evidence for game and market experiments is contradictory (Camerer and Hogarth, 1999, vs. Smith and Walker, 1993). Third, if beneficial effects are obtained, they seem to be twofold: While financial incentives do not guarantee optimal decisions, in many cases they bring decisions closer to the predictions of the normative models. Equally important, they seem to reduce error variance substantially. Fourth, all these conclusions are based on relatively small (and partly opportunistic) samples of empirical studies and, thus, may not be the last word.

This state of affairs suggests the following policy recommendation. We propose that in research areas in which performance criteria are available and in which researchers seek and intend to make inferences about maximal performance, psychologists and economists ought to make a decision about appropriate incentives. In our view, this decision should be informed by the empirical evidence available. If there is evidence in the 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 consistently 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 (Hertwig and Ortmann, 2001)—that financial incentives (or, for that matter, other key variables of experimental design) be accorded the status of independent variables in experiments. 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 the key variables of experimental design. In addition, such a procedure would distribute the effort of studying the impact of design variables among many researchers in the scientific community, thus also increasing the credibility of the obtained evidence.

In our view, a 'do-it-both-ways' policy also takes into account psychologists' concerns that obligatory financial incentives unduly stress monetary motivators at the expense of others. The systematic comparison of incentive (be they monetary or of any other nature) and non-incentive conditions would allow researchers to explore systematically the impact of different motives, for every experiment that employs financial incentives also implicitly suggests something about other motivators (e.g., altruism, trust, reciprocity, or fairness). If, for example, in the Prisoner's Dilemma games (or public good, trust, ultimatum, or dictator games), the behavior of participants does not correspond to the game-theoretic predictions—if they show more altruism (trust, reciprocity, or fairness) than the theory predicts—then these findings also reveal information about the other non-monetary motivators (assuming that the demand effects are carefully controlled, and the experiments successfully implement the game-theoretic model).

We agree with the argument that the implementation of financial incentives has its own ethical risks. But there are ways to reduce them. It is important, for example, that payments are given privately, as is the standard practice in economics experiments. We also agree that 'asking purely hypothetical questions is inexpensive, fast and convenient' (Thaler, 1987, p. 120). However, we conclude by suggesting that the benefits of being able to run many studies do not outweigh the costs of generating results of questionable reliability (see also Beattie and Loomes, 1997, p. 166).

## 6. DECEPTION: TO WHAT EXTENT DO ECONOMISTS' AND PSYCHOLOGISTS' EXPERIMENTAL PRACTICES DIVERGE, AND WHY?

Deceiving participants is generally taboo among experimental economists (Davis and Holt, 1993, p. 24); indeed, economics studies that the use of deception can probably be counted on two hands. Table 13.1 lists the statements of various prominent experimental economists who oppose deception (for a rare dissenting view in economics, see Bonetti, 1998; but see also Hey, 1998; McDaniel and



Table 13.1. Economists' rationales for prohibiting deception

---

Hey (1991, pp. 21, 119, 173, 225). I feel that it is crucially important that economics experiments actually do what they say they do and that subjects believe this. I would not like to see experiments in economics degenerate to the state witnessed in some areas of experimental psychology where it is common knowledge that the experimenters say one thing and do another. (Subjects) believing what the experimenters tells them . . . seems to me to be of paramount importance: once subjects start to distrust the experimenter, then the tight control that is needed is lost.

Davis and Holt (1993, pp. 23–24). 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.

Ledyard (1995, p. 134). It is believed by many undergraduates that psychologists are intentionally deceptive in most experiments. If undergraduates believe the same about economics, we have lost control. It is for this reason that modern experimental economists have been carefully nurturing a reputation for absolute honesty in *all* their experiments.

---

Starmer, 1998). What commonly underlies economists' opposition to deception is the fear that participants' expectations of being deceived produces suspicion and second-guessing, and that these reactions rather than the experimental scenario, instructions, and incentives guide, motivate, and ultimately distort experimental behavior.

Two mechanisms are assumed to induce suspicion and second-guessing. The first mechanism is *firsthand* experience with deception gained by participating and being debriefed in deception experiments. The second mechanism is *vicarious* experience with deception gained via communication channels such as campus scuttlebutt, media coverage of psychological research, and undergraduate teaching. Economists do assume that vicarious experiences suffice to engender the contamination of the participant pool. Therefore, they consider participants' expectation that they will not be deceived (i.e., honesty on the part of the experimenter) a common good of sorts (such as air or water) that would be depleted (contaminated), quickly even, if only a few of their tribe practiced deception.

Unlike economists, psychologists use deception. To estimate the frequency of deception, we focused on the high-ranked *Journal of Personality and Social Psychology (JPSP)* (and its predecessor, the *Journal of Abnormal and Social Psychology*), because in this journal the most comprehensive and recent figures are available. After a sharp upswing during the 1960s when the percentage of deception studies tripled from 16 percent in 1961 to 47 percent in 1971, the use of deception maintained a high level throughout the 1970s, reaching a height of

59 percent in 1979, before it dropped to 50 percent in 1983 (Adair et al., 1985). Since then it has fluctuated between 31 percent and 47 percent (1986: 32 percent; 1992: 47 percent; 1994: 31 percent; 1996: 42 percent; as reported in Adair et al., 1985; Sieber et al., 1995; Nicks et al., 1997; and Epley and Huff, 1998). Some of the fluctuations may reflect substantial changes in the applied methods (e.g., the initial upswing in the 1960s), ethical standards, and the federal regulation of research; others may reflect the 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.).

Although the use of deception has declined since its heyday in the late 1960s and 1970s, the absolute level is still high: A conservative estimate is that every third study published in *JPSP* in the 1990s employed it (compared to 4.7 percent 1921–48). In a few other social psychology journals, such as the *Journal of Experimental Social Psychology*, the proportion of deception studies appears to be even higher than in *JPSP* (e.g., Gross and Flemming, 1982; Nicks et al., 1997). Moreover, deception is not confined to research practices in social psychology—it is, for instance, also used in marketing and consumer research (see Toy et al., 1989), in personality research (see Nicks et al., 1997), and in research on behavioral decision-making (Hertwig and Ortmann, 2001).

Why do psychologists use deception? The primary motive seems to rest on two methodological arguments: First, if participants were aware of the true purpose of a study—especially when it concerns ‘sensitive’ issues (e.g., conformity, prejudices, anti-social behavior)—they might respond strategically and not reveal their true preferences, opinions, attitudes, etc.; and the investigator might lose experimental control. For instance, one might expect participants to alter their behavior in order to prove how accepting they are of members of other races if they know that they are participating in a study of racial prejudices. Therefore, so the argument goes, investigators sometimes need to camouflage the purpose of the experiment to achieve experimental control—according to Herrera (1997), this is the standard justification for using deception (see also Kelman, 1967, p. 6). Second, deception can be used to produce situations of special interest that are unlikely to arise otherwise—for instance, an emergency situation in which bystander effects can be studied. For other arguments in the support of deception, see the recent debate in the *American Psychologist* (Bröder, 1998; Kimmel, 1998; Korn, 1998; Ortmann and Hertwig, 1997, 1998).

Although the frequent use of deception appears to imply a widespread consensus among psychologists that it is a methodological necessity, there has been a longstanding and persistent concern among some psychologists regarding its long-term consequences. Table 13.2 compiles some statements from the 1960s (but similar concerns have persisted over the years, see, e.g., Wallsten, 1982). It is striking how much psychologists’ statements from the 1960s mirror economists’ statements from the 1990s. Similar to the later concerns of economists, psychologists worried that firsthand (Seeman, 1969, in Table 13.2) but also vicarious experience with deception (e.g., Adelson, 1969 and Orne, 1962, in Table 13.2)

Table 13.2. *A sample of conclusions from psychologists regarding the negative effects of deception*

---

*Orne (1962, pp. 778-79).* (The use of deception) on the part of psychologists is so widely known in the college population that even if a psychologist is honest with the subject, more often than not he will be distrusted. As one subject pithily put it, "Psychologists always lie!" This bit of paranoia has some support in reality.

*Ring (1967, p. 118).* What is the perceptive student to think, finally, of a field where the most renowned researchers apparently get their kicks from practicing sometimes unnecessary and frequently crass deceptions on their unsuspecting subjects? . . .

The short-run gains may be considerable, but it does not appear chimerical to suggest that the ultimate price of deception experiments may be the creation of extremely mistrustful and hostile subject pools. It would be ironic indeed if, by their very style of research, social psychologists were to put themselves out of business.

*Kelman (1967, p. 6).* How long, however, will it be possible for us to find naïve subjects? Among college students, it is already very difficult. They may not know the exact purpose of the particular experiment in which they are participating, but at least they know, typically, that it is *not* what the experimenter says it is. . . . If he resents the experimenter's attempt to deceive him, he may try to throw a monkey wrench into the works; I would not be surprised if this kind of Schweikian game among subjects became a fairly well-established part of the culture of sophisticated campuses.

*Argyris (1968, p. 187).* Many experiments have been reported where it was crucial to deceive the students. . . . One result that has occurred is that students now come to experiments expecting to be tricked. The initial romance and challenge of being subjects has left them and they are now beginning to behave like lower level employees in companies. Their big challenge is to guess the deception (beat the management). If one likes the experimenter, then he cooperates. If he does not, he may enjoy botching the works with such great skill that the experimenter is not aware of this behavior.

*Adelson (1969, p. 220).* . . . when the campus population learns, as it can hardly fail to do, about the common tendency of psychologists to deceive, so that all kinds of unanticipated, unknown expectations enter the experimental situation, the subject aiming to 'psych' the experimenter's 'psyching' of him, subject and experimenter entangled in a web of mutual suspicion, mutual deception.

*Seeman (1969, pp. 1025-6).* When a subject has once participated in a study using deception he is no longer a naïve subject but a sophisticated subject who brings to subsequent studies a variety of personal theories and hypotheses that guide the behavior of the subject quite as decisively as theories and hypotheses guide the behavior of an experimenter. In view of the frequency with which deception is used in research we may soon be reaching a point where we no longer have naïve subjects, but only naïve experimenters. It is an ironic fact that the use of deception, which is intended to control the experimental environment, may serve only to contaminate it.

---

would create the expectation among participants that they will be tricked in experiments, making them distrustful or even hostile. A discipline encumbered with the reputation of using deception, it was argued, would compromise the very asset deception was meant to secure—experimental control (see e.g., Ring, 1967, and Seeman, 1969, in Table 13.2).

## 7. DECEPTION: WHAT ARE THE CONSEQUENCES OF ITS USE?

Is there evidence that firsthand, or even just vicarious, experiences with deception destroys experimental control? Moreover, is there any evidence that participants' suspicion—through whatever experiences it is brought about—has negative consequences? Drawing on our analyses (Hertwig and Ortmann, 2002; Ortmann and Hertwig, in press), we report some of the major empirical findings. We begin with the second question.

### 7.1. Is Suspicion an Inconsequential Side Effect of Deception?

Could suspicion *per se* ultimately be inconsequential because the experimental situation is real—or innocuous—enough? There are at least three ways to examine this question. One way is to record the participants' suspicion post-experimentally (before debriefing them) and then analyze people's performance as a function of it. A second way is to take the bull by the horn, engendering participants' suspicion from the outset and studying their subsequent performance as a function of it. Yet another approach considers psychologists' institutional arrangements and their evolution over time as an indication of their regard toward suspicion and its potentially damaging influence on experiments (see Ortmann and Hertwig, in press).

#### 7.1.1. *Post-Experimental Identification of Suspicion*

To find relevant studies, we (Hertwig and Ortmann, 2002) searched the PsycINFO database using the keyword 'deception' in combination with 'suspicion'. Our search turned up 14 studies that examined experimental performance as a function of suspicion. In toto, this collection can be taken as a case study of the effects of suspicion and distrust on a dependent variable under consideration. All of these studies were concerned with conformity behavior—a research topic that arose out of Solomon Asch's remarkable finding that people with a normal vision would ignore what they had seen, in order to agree publicly with an obviously inaccurate group judgment.

In Asch's paradigm, participants are recruited to participate in a visual discrimination task. They are asked to announce publicly which one of three comparison lines matched a standard length. Although seven to nine people typically participate in each session, only one is a naïve participant: The others are instructed to answer correctly on the first two trials, then incorrectly but unanimously on the remaining trials (see Cialdini and Trost, 1998).

Do participants who suspect whether experimenter is speaking the truth—not fully at least—exhibit the same kind of conformity behavior as non-suspicious participants? Among the 14 studies we analyzed (Hertwig and Ortmann, 2002), we found the following: In ten studies (71 percent), suspicious participants conformed less than non-suspicious participants. In nine of these, enough information was given to calculate the effect size  $\eta$  (see Cohen, 1988), the reduction in conformity due to suspicion was of a medium to large effect size (for details see table 3 in Hertwig and Ortmann, 2002). In the remaining four studies, suspicion did not significantly change the amount of conformity behavior. Albeit limited to the conformity paradigm, these results demonstrate that participants' suspicion *can* systematically alter the very behavior the experimenter intends to measure.

A final remark on conformity studies: In an extensive search of the conformity literature, Stang (1976) identified 21 studies that reported the percentage of participants who were classified as suspicious. Interestingly, he found that participants' suspicion has risen through the decades, with particularly steep increases in the second half of the 1960s (which corresponds closely to the dramatic increase of deception experiments during that period).

#### 7.1.2. *Ex-Ante Manipulation of Suspicion*

A small group of studies actively planted suspicion in order to examine its effects (e.g., participants were given detailed tip-offs about the true purpose of the experiment). Some studies found no effect from this manipulation, and others a good deal. Nevertheless a trend is discernable (for detailed results see table 4 in Hertwig and Ortmann, 2002). When participants receive specific, definite information about deception, experimental performance is indeed altered. However, if the information is vague and deception is merely a possibility, performance is not likely to change (in comparison to a control group).

Is this reason enough to call off the alarm? Not according to Finney (1987, p. 45), who argued that being informed about the possibility of deception 'merely reaffirms subjects' prior belief that deception may occur in an experiment and, therefore, causes no change in their anticipation' (in comparison to participants in the control group who share the same prior belief).

#### 7.1.3. *Psychologists' Institutional Arrangements*

If psychologists believed suspicion to be inconsequential, they probably would not bother to take any measures against it. But they appear to do so—both on an individual and collective level. For instance, a colleague of ours who attended the CEPR conference on psychology and economics told us that at his laboratory, in which deception is used and in which students are eligible to participate in multiple experiments, the experimenter routinely probes for suspicion at the end of the studies. In addition, the experimenter asks the participants to list the previous studies they have participated in. If the experimenter needs naïve

participants, she can discount all data from participants who have previously participated in a deception study. Or she might choose to analyze those data separately and estimate the effects of experience with prior deceptions. As far as we know, this kind of arrangement is not institutionalized throughout psychological laboratories but are left to the discretion of the individual researcher (and is thus likely to be contingent on his or her access to participants as well as financial resources).<sup>2</sup>

Beyond the obvious ways to deal with the possible consequence of participants' suspicion, there also appear to be more subtle arrangements. Consider, for instance, the salient changes in the selection and composition of psychology's subject pool. Sieber and Saks (1989) reported the responses of 326 psychology departments with subject pools (see also Vitelli, 1988). They found that of the 74 percent that reported having a participant pool, 93 percent recruited from introductory courses. In contrast, in his summary of human participant sources in three journals of the American Psychological Association, Schultz (1969, table 1, p. 217) found, on average, that fewer than 40 percent of human participants were recruited from introductory psychology courses. While the data are not completely comparable, they suggest that during the two decades between Schultz (1969) and Sieber and Saks (1989), the percentage of participants from introductory courses has roughly doubled.

One speculation is that the peculiar institutional arrangements in psychology (namely, the widespread use of participants from introductory courses) result from an evolutionary process: psychology departments have increased their attempts to minimize participants' suspicion by relying on participants who are less likely to have a firsthand experience with deception. In other words, one way to read the two snapshots presented in Schultz (1969) and Sieber and Saks (1989) is that psychologists took the advice of Silverman et al. (1970, p. 211) 'that the practice of using the same subjects repeatedly be curtailed, and whenever administratively possible, subjects who have been deceived and debriefed be excluded from further participation.'

In our view, the available evidence suggests that suspicion is not simply an inconsequential side effect of deception but has the potential to alter the very behavior the experimenter intends to measure. Judging from the evolution of various institutional arrangements, psychologists seem to assume it does. How is participants' suspicion caused in the first place? Does it require firsthand experience (which was present in the above mentioned studies on the impact of suspicion on conformity behavior or on the impact of planted suspicion) or does vicarious experience suffice?

<sup>2</sup> The failure to standardize such arrangements may have the unfortunate, and paradoxical, consequence that researchers who do not use deception are more likely to become a victim of its potentially distorting effects: they might be less inclined to probe their participants for suspicion and, thus, be less able to control for the effect of prior experience with deception.

## 7.2. Is Vicarious Experience as Consequential as Firsthand Experience?

Unfortunately, there is scant empirical evidence to answer this question. We have found very few attempts to quantify participants' vicarious experience. The little evidence available (summarized in Hertwig and Ortmann, 2002) is mostly concerned with the effect of undergraduate teaching. According to Rubin and Moore (1971), undergraduate training, specifically the teaching of classic social psychological experiments (e.g., Milgram experiments, conformity experiments), can serve to develop students' expectation of what experimenters do. They observed that the number of psychology courses which students had completed—not the number of deception experiments in which participants recall having taken part—correlated with the participants' level of suspicion. In line with this result, Higbee (1978) observed that students rated psychologists as being less truthful at the end of the semester than at the beginning ( $\eta = 0.51$ ), and students who had taken at least five psychology courses rated psychologists as less truthful than students without any course experience ( $\eta = 0.43$ ).

Higbee concluded 'if psychologists expect the subjects to believe them, perhaps they should get the subjects at the beginning of the semester' (Higbee, 1978, p. 133; a refinement of the advice given by Silverman et al., 1970). The widespread use of students from introductory courses (described above) could be an attempt to approximate this advice.

## 8. DECEPTION: WHERE TO GO FROM HERE?

The methodological costs of deception are neither well explored nor well understood. The little we know, however, does help us to evaluate both the present experimental practices and calls to change them. We believe that the available evidence is not sufficient to convince researchers in psychology to abandon a widely used and powerful research tool. In particular, the lack of research regarding the effects of vicarious experience does not allow substantiation of the reputational spillover effects predicted by various researchers (see Tables 13.1 and 13.2). We caution, however, that such a lack may be the result of institutional responses to the very problem. That said, the evidence regarding the consequences of firsthand experience with deception counsels us to treat deception as a last-resort strategy, thus minimizing the number of participants with firsthand experience. This is in fact the policy currently recommended in the guidelines of the American Psychological Association (APA).<sup>3</sup>

<sup>3</sup> According to the APA rules of conduct, '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). In other words, the APA rules allow that deception may be indispensable under certain circumstances but treat it as a last-resort strategy, to be used only if its benefits justify its use *and* there are no feasible alternatives.

Even a cursory glance at contemporary deception studies, however, reveals that deception is not treated as a last resort. Consider the following study—recently published in a social psychological journal—which had participants play, among others, an ultimatum game. Participants were falsely told that they would be paired with one of the other participants, that on the basis of a chance procedure they alone were assigned the role of the allocator (who had to divide a certain amount of money), and that their income would be contingent on the allocator's and/or responder's decisions. In addition, the participants were told that the experimenter would hand the allocator's decision to the recipient (involving a rather complicated procedure) and thus decisions would remain anonymous. At the end of the experiment, participants were debriefed, discovering that all participants were allocators and that all received the same amount of money.

This study is not an example of dramatic deception. In fact, the lies were rather mild. This study, however, is a good example of the use of deception where it is completely unnecessary (as witnessed by the many economic studies researching ultimatum games without deception). In the present case, deception was probably used to increase the number of players who are allocators. Sometimes, deception and the lack of financial incentives provide the same benefits: Both practices are inexpensive, fast, and convenient. Deception, however, is only inexpensive if there will be no costs for future experiments. Indeed, why would these participants, who ultimately found out that their decisions and monetary rewards were not contingent, believe the promises of performance-dependent monetary rewards in future experiments? Mistrust of such a promise in future experiments has the clear potential to affect their experimental performance.

This study demonstrates, along with evidence of its widespread use in psychology, that deception is an accepted way of doing business rather than a strategy of last resort, as recommended by the APA. Although the APA rules of conduct are considerably stricter now than in the 1960s (and indeed seem to have successfully reduced the severity of deceptive techniques), they have not changed (some) psychologists' somewhat cavalier approach to deception.

The question is, why—even after the very public debates of the 1970s—are the APA rules of conduct not effective in enforcing deception as a last-resort strategy? Elsewhere we have argued that the main problem is that the key decision whether deception is justified by its anticipated utility is left to those that stand to benefit from its use (Ortmann and Hertwig, 1997, 1998). Notwithstanding the mediating role of institutional review boards (which tend to focus on the ethical rather than the methodological consequences of deception), this practice leaves the assessment of private benefits (e.g., relatively quick publication, see Adelson, 1969) and public costs (contamination of the participant pool) to the interested party (the experimenter)—a classic moral-hazard problem with a solution that is currently not incentive-compatible.

We have suggested a pragmatic solution that we believe is significantly more incentive-compatible (Hertwig and Ortmann, 2002). Specifically, we have proposed that experimenters about to perform deception studies post their



experimental designs on an APA website for a specified time period, thus giving those opposed to deception a chance to suggest workable alternatives. Such a procedure might spur a spirited case-by-case debate about deception's necessity, and might ultimately lead to methodological innovation. Over time, such a website would offer successful alternatives, with examples of experiments in which they were used, so that experimenters considering deception could easily 'browse' through them.

We are aware that the evidence and the conclusions expressed in this chapter are not undisputed. In either field there are researchers who disagree with our views, while others are more favorable. For an extensive sample of opinions see the comments on our target article in the *Behavioral and Brain Sciences* (Hertwig and Ortmann, 2001).

## 9. EPILOGUE

To look at methodological practices across disciplinary borders can be extremely useful, as it allows us to put our daily routines, habits, and beliefs into perspective and to reflect on their costs and benefits. Economics and psychology—two related disciplines—have surprisingly different conceptions of good experimentation even in highly overlapping areas such as research on behavioral decision-making. We believe that both disciplines should use their pronounced differences as an opportunity to more closely and systematically evaluate their methodological preferences. We hope that this chapter will contribute to a spirited discussion of the costs and benefits of those preferences.

## REFERENCES

- Adair, J.G., Dushenko, T.W., and Lindsay, R.C. L. (1985). Ethical regulations and their impact on research practice. *American Psychologist*, 40, 59–72.
- Adelson, J. (1969). Personality. *Annual Review of Psychology*, 20, 217–52.
- American Psychological Association (1992). Ethical principles of psychologists and code of conduct. *American Psychologist*, 47, 1597–611.
- Argyris, C. (1968). Some unintended consequences of rigorous research. *Psychological Bulletin*, 70, 185–97.
- Beattie, J. and Loomes, G. (1997). The impact of incentives upon risky choice experiments. *Journal of Risk and Uncertainty*, 14, 155–68.
- Bonetti, S. (1998). Experimental economics and deception. *Journal of Economic Psychology*, 19, 377–95.
- Bröder, A. (1998). Deception can be acceptable. *American Psychologist*, 53, 805–6.
- Camerer, C.F. (1990). Do markets correct biases in probability judgment? Evidence from market experiments. In L. Green and J. Kagel (Eds.), *Advances in behavioral economics* (pp. 125–72). Northwood, NJ: Ablex Publisher.
- (1995). Individual decision-making. In J.H. Kagel and A.E. Roth (Eds.), *Handbook of experimental economics* (pp. 587–703). Princeton: Princeton University Press.

- and 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.
- Chu, Y.-P. and Chu, R.-L. (1990). The subsidence of preference reversals in simplified and marketlike experimental settings: A note. *American Economic Review*, 80, 902–11.
- Cialdini, R.B. and Trost, M.R. (1998). Social influence: Social norms, conformity, and compliance. In D.T. Gilbert, S.T. Fiske and G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 1, 4th ed.). Boston: McGraw-Hill.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.
- Davis, D.D. and Holt, C.A. (1993). *Experimental economics*. Princeton: Princeton University Press.
- Dawes, R.M. (1996). The purpose of experiments: Ecological validity versus comparing hypotheses. *Behavioral and Brain Sciences*, 19, 20.
- (1999). Experimental demand, clear incentives, both, or neither? In D. V. Budescu, I. Erev, and R. Zwick (Eds.), *Games and human behavior: Essays in honor of Amnon Rapoport* (pp. 21–8). Mahwah, NJ: Erlbaum.
- Deci, E.L., Koestner, R., and Ryan, R.M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125, 627–68.
- Eisenberger, R. and Cameron, J. (1996). Detrimental effects of reward. Reality or myth? *American Psychologist*, 51, 1153–66.
- Epley, N. and 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.
- Finney, P.D. (1987). When consent information refers to risk and deception: Implications for social research. *Journal of Social Behavior and Personality*, 2, 37–48.
- Grether, D.M. (1978). Recent psychological studies of behavior under uncertainty. *American Economic Review*, 68, 70–4.
- and Plott, C.R. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review*, 69, 623–38.
- Gross, A.E. and Fleming, I. (1982). Twenty years of deception in social psychology. *Personality and Social Psychology Bulletin*, 8, 402–8.
- Harrison, G.W. (1989). Theory and misbehavior of first-price auctions. *American Economic Review*, 79, 749–62.
- (1992). Theory and misbehavior of first-price auctions: Reply. *American Economic Review*, 82, 1426–43.
- (1994). Expected utility theory and the experimentalists. *Empirical Economics*, 19, 223–53.
- Herrera, C.D. (1997). A historical interpretation of deceptive experiments in American psychology. *History of the Human Sciences*, 10, 23–36.
- Hertwig, R. and Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences*, 24, 383–451.
- and — (2002). Does deception destroy experimental control? A Review of the evidence. Manuscript submitted for publication.
- Hey, J.D. (1991). *Experiments in economics*. Oxford: Basil Blackwell.
- (1998). Experimental economics and deception: A comment. *Journal of Economic Psychology*, 19, 397–401.

- Higbee, K.L. (1978). How credible are psychological researchers to college students? *The Journal of Psychology*, 99, 129-33.
- Hogarth, R.M. and Reder, M.W. (Eds.) (1987). *Rational choice: The contrast between economics and psychology*. Chicago: University of Chicago Press.
- Jenkins, Jr., G.D., Mitra, A., Gupta, N., and Shaw, J.D. (1998). Are financial incentives related to performance? A meta-analytic review of empirical research. *Journal of Applied Psychology*, 83, 777-87.
- Kelman, H.C. (1967). Human use of human subjects: The problem of deception in social psychological experiments. *Psychological Bulletin*, 67, 1-11.
- Kimmel, A.J. (1998). In defense of deception. *American Psychologist*, 53, 803-5.
- Koehler, J.J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, 19, 1-53.
- Korn, J.H. (1998). The reality of deception. *American Psychologist*, 53, 805.
- Ledyard, J.O. (1995). Public goods: A survey of experimental research. In J.H. Kagel and A.E. Roth (Eds.), *Handbook of Experimental Economics* (pp. 111-94). Princeton: Princeton University Press.
- McDaniel, T. and Starmer, C. (1998). Experimental economics and deception: A comment. *Journal of Economic Psychology*, 19, 403-9.
- Nicks, S.D., Korn, J.H., and Mainieri, T. (1997). The rise and fall of deception in social psychology and personality research, 1921 to 1994. *Ethics and Behavior*, 7, 69-77.
- 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.
- Ortmann, A. and Hertwig, R. (1997). Is deception acceptable? *American Psychologist*, 52, 746-7.
- and ——— (1998). The question remains: Is deception acceptable? *American Psychologist*, 53, 806-7.
- and ——— (in press). The costs of deception: Evidence from psychology. *Experimental Economics*.
- Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, 37, 7-63.
- Rabin, M. (1998). Psychology and economics. *Journal of Economic Literature*, 36, 11-46.
- Ring, K. (1967). Experimental social psychology: Some sober questions about some frivolous values. *Journal of Experimental Social Psychology*, 3, 113-23.
- Rubin, Z. and Moore, Jr., J.C. (1971). Assessment of subjects' suspicions. *Journal of Personality and Social Psychology*, 17, 163-70.
- Schultz, D.P. (1969). The human subject in psychological research. *Psychological Bulletin*, 72, 214-28.
- Seeman, J. (1969). Deception in psychological research. *American Psychologist*, 24, 1025-8.
- Sieber, J.E., Iannuzzo, R., and Rodriguez, B. (1995). Deception methods in psychology: Have they changed in 23 years? *Ethics and Behavior*, 5, 67-85.
- and Saks, M.J. (1989). A census of subject pool characteristics and policies. *American Psychologist*, 44, 1053-61.
- Silverman, I., Shulman, A.D., and Wiesenthal, D.L. (1970). Effects of deceiving and debriefing psychological subjects on performance in later experiments. *Journal of Personality and Social Psychology*, 14, 203-12.

- Smith, V.L. (1976). Experimental economics: Induced value theory. *American Economic Review*, 66, 274–9.
- (1982). Microeconomic systems as an experimental science. *American Economic Review*, 72, 923–55.
- Smith, V.L. (1991). Rational choice: The contrast between economics and psychology. *Journal of Political Economy*, 99, 877–97.
- and Walker, J.M. (1993). Monetary rewards and decision costs in experimental economics. *Economic Inquiry*, 31, 245–61.
- Stang, D.J. (1976). Ineffective deception in conformity research: Some causes and consequences. *European Journal of Social Psychology*, 6, 353–67.
- Thaler, R. (1987). The psychology of choice and the assumptions of economics. In A. E. Roth (Ed.), *Laboratory experimentation in economics: Six points of view* (pp. 99–130). Cambridge: Cambridge University Press.
- Toy, D., Olsen, J., and Wright, L. (1989). Effects of debriefing in marketing research involving 'mild' deceptions. *Psychology & Marketing*, 6, 69–85.
- Tversky, A. and Kahneman, D. (1987). Rational choice and the framing of decisions. In R. M. Hogarth and M.W. Reder (Eds.), *Rational choice: The contrast between economics and psychology* (pp. 67–94). Chicago: University of Chicago Press.
- Vitelli, R. (1988). The crisis issue assessed: An empirical analysis. *Basic and Applied Social Psychology*, 9, 301–9.
- Wallsten, T.S. (1982). *Research through deception* (Letter to the editor). New York Times Magazine, October.