When Economic Experiments Can Help Consumer Research

Steven C. Huff [1]

[1] Woodbury School of Business, Utah Valley University, USA

Correspondence: Steven C. Huff, 800 W. University Pkwy-MS 288, Orem, Utah 84058, USA. Tel: 1-801-863-8863. E-mail: huff@uvu.edu, professorhuff@gmail.com

Abstract

In their quest to better understand and predict human behavior, consumer researchers have been relatively unaffected by a growing trend in the social sciences—the interplay between economics and psychology. This is not necessarily a surprise given that consumer researchers are often trained within the psychology paradigm and often get little exposure to economics. The main purpose of this article, therefore, is to (a) introduce consumer researchers to methods of experimental economics, and (2) propose situations in which its tools and procedures can provide alternative approaches to common problems in consumer research. The article is divided into two main sections. First, it describes four key practices that characterize a typical economic experiment, and discusses how they can be applied to consumer research. The article then presents two procedures commonly used in experimental economics to elicit preferences and beliefs.

Keywords: experimental economics, consumer research, lab experiments, methodology

1. Introduction

Experimentation is a common methodology used in consumer research. However, most of the experiments found in consumer-focused journals follow the experimental norms used in psychology. In contrast, the experimental methods developed in economics are virtually absent.

One might suggest that experimental economics has little to offer to consumer research; hence, its absence. A more likely explanation, however, arises from the traditional dichotomy in marketing as a field—i.e., quantitative/economics training vs. consumer behavior/psychology training. Experimental economics and consumer behavior not only adopt similar methodological approaches (i.e., lab and field experiments), but they also share similar goals—i.e., to better understand and predict how individuals behave in the market place. Consumer research's inherent interdisciplinary nature would benefit significantly from a trend that is already in course in the social sciences—i.e., the interplay between psychology and economics.

The main purpose of this article is to (a) introduce psychologically-trained consumer researchers to experimental economics, and (b) propose situations in which its tools and procedures can provide alternative approaches to common problems in consumer research. It is not the purpose of this article to promote experimental economics as a superior or replacement methodology for traditional psychological experiments, but rather to discuss when the practices of experimental economics can augment current practices in consumer research. It is also not the purpose of this article to answer the legitimate criticisms of experimental economics. As with all research practices, the experimental economics approach has strengths and weaknesses, and while some of these issues are mentioned here, they are not discussed in detail. This article has two main sections. First, it describes the four key practices that characterize a typical economic experiment, and discusses how each applies to consumer research. Then, the article presents two common procedures used in experimental economics to elicit preferences and beliefs. Throughout these two sections, suggestions for future research are made. Finally, the article concludes.

2. Four Key Practices in Economic Experiments

An economic experiment commonly includes four basic practices that distinguish it from a typical psychological experiment: abstraction, stationary replication, consequential incentives, and the absence of deception. In this section, each practice will be described and then the article will discuss how it may be applied to consumer research.

2.1 Abstraction

2.1.1 Explanation

Abstraction in economic experiments refers to the practice of employing general and neutral terminology in experimental instructions and stimulus materials (Davis & Holt, 1993; Holt, 1995; Kachelmeier & Shehata, 1997). Economists believe that abstraction provides three important benefits: generalization of experimental results (i.e., greater external validity), better theory testing, and better replication.

2.1.2 Generalizability

Suggestive terms can alter experimental results significantly. For example, Ross and Ward (1996) and Liberman, Samuels, and Ross (2004) report that changing the title of a simple prisoner's dilemma game from "The Community Game" to the "Wall Street Game" drops the rate of cooperation significantly. Burnham, McCabe, and Smith (2000) make a more subtle change to a two-player reciprocity game. They simply change the work "opponent" in the experimental instructions to "partner" and find that this small change results in more than a doubling of trust as measured by experimental outcomes.

Abstracting experimental materials through the use of relatively neutral terms represents an attempt to increase the generalizability of experimental results by reducing the biases due to the use of suggestive terms (e.g., Fehr & Zych, 1998; Fehr, Gächter, & Kirchsteiger, 1997; Güth, Ivanova-Stenzel, & Tjotta, 2004; Irlenbusch & Sutter, 2006; Potters & van Winden, 2000). For example, the following instruction describes an auction experiment:

Today we are going to set up a market in which some of you will be buyers and others will be sellers. The commodity to be traded is divided into distinct items or 'units'. We will not specify a name for the commodity; we will simply refer to units. Trading will occur in a sequence of trading periods. The prices that you negotiate in each trading period will determine your earnings, in dollars and cents (Davis & Holt, 1993).

This instruction employs relatively neutral terms (e.g., buyers, sellers, market, units, etc.) to describe the roles participants will take on and the context in which decisions will be made. It is sufficiently abstract in that it is unlikely to invoke behaviors associated with a specific market setting (e.g., like an auction) or a specific type of buyer or seller in the market. Contrast this with an alternative version of the same instruction:

Today we are going to set up an auction in which some of you will be art collectors and others will be art auctioneers. The product to be auctioned is paintings. The experiment will consist of a sequence of auctions. The prices that you negotiate through bidding in each auction will determine your wealth, in dollars and cents.

This instruction uses terminology that is specific to an art auction. It is less abstract because the specific terms employed in the instruction (e.g., art collectors, auctioneers, auction, paintings, and bidding) may induce participants to behave in a manner they believe to be proper at an art auction, which may differ significantly from the behavior participants believe to be appropriate in other auction environments (i.e., e-bay, cattle auctions, etc.). Of course, if the explicit goal of the researcher is to study art auctions, this can be desirable, but even in this case it can be instructive to run multiple conditions with specific and generalized terminology to discover how this specificity alters consumer behavior.

2.1.3 Theory Testing

The temptation to design initial experiments to mimic the real world can be strong, but if the purpose of an experiment is to test theory, then the goal should not necessarily be to recreate the real world. Instead, researchers might want to employ the most parsimonious procedure possible that still contains the key elements needed to test the theory of interest. Such procedures can be so simple that they bear little resemblance to the real world situations to which the theory applies. Abstract terminology can help simplify the experimental context and therefore make the conclusions of the theory easier to observe and test. The strength of this approach is that if even in its simplest form, the theory does not hold, then its ability to hold in the real world seems less plausible. When the researcher wishes to "stress test" the theory, it then becomes appropriate to add increasing real-world complexity to experimental approaches.

2.1.4 Replication

Abstraction can also aid researchers in the replication and comparison of results. If a researcher wishes to compare how a market in a retail setting (where sellers set prices) compares to a market in an auction setting (where buyers bid prices), abstraction allows the experimental designs to be very similar—differing in a few moderating variables of interest (e.g., whether the price is proposed by the buyer or seller), but employing the same terminology and measuring the same dependent variables. The cumulative result of this approach is a large quantity of easily comparable experiments, where the robustness of the findings can be readily assessed

(Camerer, 2003). This is one of the great strengths of the economic approach to experimentation.

2.1.5 Abstraction in Consumer Research

Few experiments in consumer research adopt abstract or relatively de-contextualized instructions. This is not necessarily a surprise. Consumer researchers have traditionally followed the opposite route, such that contextualization, rather than abstraction, represents one of the key ingredients of a proper experiment, because rich detail should increase external validity. However, external validity associated with a single real-world context is not the same thing as generalizability to many contexts. In order to achieve generalizability, some consumer researchers have applied variations of an experimental procedure to multiple contexts (e.g., five different product categories). This practice should be preferred when the nature of dependent variables are such that abstraction is not feasible. However, when a relatively context-neutral variation of the experimental design can be performed (and this can be in addition to experiments with more specific terminology), the results can significantly strengthen the claim to generalizability. Moreover, consumer researchers might be interested in using "level of abstraction" as a moderating factor, where the highly abstract condition represents a "baseline," and more contextualized scenarios address additional aspects of interest.

2.2 Stationary Replication

2.2.1 Explanation

Experimental economists predominantly study what they call steady-state or equilibrium behavior (Camerer, 1997). In other words, economists favor knowing what decision a person will typically make after facing the decision many times over knowing what that person will decide when facing the decision the first or second time. In the lab, economists worry that initial behaviors observed during an experimental session may not persist as participants get more comfortable with the lab environment and as they further explore the choices available to them (Binmore, 1994). Furthermore, when participants make decisions that interact with the decisions of other participants (as in a market or game experiment—discussed later), their decisions may change as they learn how to maximize their payoffs within such interaction (Hertwig & Ortmann, 2001).

For this reason, experimental economists, whenever possible, will conduct repeated trials (i.e., decision rounds) with the participants repeatedly making the same decisions under the same conditions in a single experimental session. This practice is called *stationary replication*. Note that in interaction experiments, a *matching protocol* is typically used to randomly assign experimental participants in each round so that no two participants ever interact twice during the experimental session. This practice eliminates reputation effects that would violate the "same conditions" requirement of stationary replication.

In addition to studying steady-state decisions, stationary replication has also been used to study how experimental trials converge toward the prediction of a theory. Such is the case in the learning literature, for instance, where the researcher is interested in how quickly participants converge to their steady-state behaviors and which factors influence the speed and nature of this convergence (see Camerer & Ho, 1999, for an example of how learning is studied in experimental economics).

2.2.2 Stationary Replication in Consumer Research

As with abstraction, stationary replication is uncommon in consumer research. However, there are many instances in which repeated trials seem appropriate in consumer contexts. First, consumer researchers care about consumer learning, and stationary replication can be useful in assessing it. For example, although cues and heuristic processing often bias consumer decisions, it is less clear the extent to which learning plays a role in the process. Hertwig and Ortmann (2001) examine the articles in Koehler's (1996) comprehensive review of Bayesian reasoning research; specifically they focus on psychology articles that contain original data demonstrating Kahneman and Tversky's (1973) base-rate fallacy. Of the 106 articles examined, only 11 (10%) provided participants with feedback after each trial. But Harrison (1994) demonstrates that when participants gained experience through feedback over repeated trials they were not influenced by the representative heuristic (the explanation given by Kahneman and Tversky for the base-rate fallacy).

Second, consumer researchers are also interested in consumer-buyer interactions. Since many of these interactions repeat themselves multiple times in a relatively short period of time (e.g., tipping or price bargaining), stationary replication seems an appropriate approach to adopt. How does behavior vary as a function of the repeated interaction? How does reputation influence repeated interactions? These are common questions that can be addressed with repeated trials.

2.3 Consequentialism

2.3.1 Explanation

Smith (1982) specifies conditions for a proper incentive design in an economic experiment. This article will focus on two: salience and dominance. An experiment is *salient* if participants' rewards are tied to their decisions or actions. Thus, an experiment that simply pays participants for participation is not salient because their decisions are independent of their payment. Put differently, a salient experiment must make payoffs dependent on choice alternatives. An experiment satisfies the condition of *dominance* if the rewards tied to decisions dominate any subjective (i.e., non-financial) costs associated with experimental participation. That is, rewards (monetary or otherwise) must be sufficiently high so that it is worthwhile for participants to apply their best effort toward accomplishing an experimental task. Note that because dominance requires that rewards be tied to decisions, if the dominance criterion is satisfied, the salience criterion is necessarily satisfied also.

From here on, incentives that meet these two criteria (salience and dominance) will be referred to as "consequential" incentives. Consequential incentives in experiments can provide important benefits. They can (a) capture emotional reactions integral to the decision making process, (b) reduce demand artifacts, (c) change learning patterns, and (d) increase statistical power and reduce type I & II errors.

2.3.2 Emotional Reactions

It is well established that people have a hard time predicting how they would behave in an actual situation (Holt & Laury, 2002; Loewenstein & Schkade, 1999). Part of the reason is that even minor emotional changes can influence judgment and decision making (for a review, see Cohen, Pham, & Andrade, 2008) and hypothetical environments tend to mitigate the intensity of the emotion and its impact on decision making. For example, Loewenstein and Adler (1995) showed that the endowment effect is mitigated when people face hypothetical instead of actual scenarios. Specifically, participants who owned the target object asked for higher selling prices than those asked to imagine they owned the object. In a similar vein, Andrade and Iyer (2009) demonstrate that in a sequence of two fair gambles, people plan to reduce the amount bet after an anticipated loss, but end up betting more than planned once the loss is *felt*. As consequential incentives are intrinsically associated with emotional reactions and people frequently misestimate the impact of emotions on their own future preferences (for a review, see Loewenstein & Schkade, 1999; Van Boven & Kane, 2006), the use of consequential incentives is under many circumstances likely to be beneficial to our understanding of how consumers behave in the market place.

2.3.3 Demand Artifacts

A common problem in experimental research refers to participants' guessing of the hypothesis being tested, and, as result, their deliberate or even unconscious attempt to confirm the researcher's intuitions. Such demand artifacts are problematic as they may lead to an inferential mistake with regard to the cause of a given observed effect (Kruglanski, 1975; Sawyer, 1975; Shimp, Hyatt, & Snyder, 1991). A common solution has usually been to identify and throw out hypothesis guessers. As Shimp et al. point out, there are several problems with this approach. First, the measurement of hypothesis guessing is far from perfect. For instance, many participants may not want to spend time elaborating on the true purpose of the experiment at the end of a tiring session. Also, many may properly guess the hypothesis only when they are prompted with the question and asked to directly think about it. In the former case, there are participants who know the hypothesis but are not motivated to properly elaborate on it in an open-ended question, whereas, in the latter case, there are participants who are in-hindsight-hypothesis-guessers. Second, those who indeed knew the hypothesis and reported it at the end of the session might or might not have been influenced by it during the experiment. In other words, guessing a researcher's hypothesis does not necessarily mean that a participant will try to (dis)confirm the researcher's hypothesis.

Consequential incentives can minimize demand artifacts by reducing the impact of hypothesis guessing. The reason is simple. When no incentives are at stake, "pleasing the experimenter" is an appealing option and thus, this is more likely to bias one's judgment and decision-making. As the consequences (financial or otherwise) of a decision become more prominent, however, participants might naturally be less concerned about matching their behavior to the researcher's hypothesis than to choose the option that best improves their wellbeing. Take Andrade and Ho's (2007) experiment on interpersonal negotiations and emotions as an example. The authors show that proposers take into consideration receivers' incidental affective state (anger vs. happiness) before making an offer in the ultimatum game. In a series of rounds where proposers were matched with different receivers, right before the offer was made, the proposer was informed on whether the receiver in that particular round was happy (had watched an episode of the sitcom "Friends") or angry (had watched a drama where an

employee reacts aggressively after being fired by an arrogant boss). It seemed clear by the design (a within-participants manipulation of receiver's incidental affect) and procedure that the experimenter wanted to know whether proposers would take receivers' incidental feelings into account before making an offer. In other words, demand artifacts were possible in such circumstances. However, since the experiment used consequential incentives, an attempt to confirm the experimenter's hypothesis would increase proposers' chances of losing part of their own participation fee (and they were fully aware of it).

In short, whereas demand artifacts are a common problem and may be present in many experiments in consumer behavior, the use of consequential incentives can in general help attenuate its effect, as it creates a tradeoff between choosing the option that best fits (a) the participant's well-being versus (b) the experimenter's hypothesis. In a classic economic experiment, participants often have a clear idea of what the experimenter wants to know (see discussion on deception latter on), however, given that consequential incentives are at stake, participants are usually more concerned about their financial wellbeing than about hypothesis guessing and/or impression management.

2.3.4 Learning

In many experimental scenarios participants have the opportunity to learn from previous decisions before making subsequent choices. There is robust evidence to suggest that consequential incentives increase learning effort (Castellan, 1969; Irwin, McClelland, & Schulze, 1992; Jamal & Sunder, 1991; List & Cherry, 2000; Phillips & Edwards, 1966). For example, List and Cherry (2000) demonstrate the influence of incentives on learning with the ultimatum game. The basic idea of this two-player game is that Player 1 proposes how to divide a sum of money and Player 2 accepts (resulting in the agreed upon division of the sum of money) or rejects (resulting in a payoff of zero for both players). The theoretical prediction for this game is that Player 1 proposes to give Player 2 the smallest, non-zero amount possible and Player 2 accepts. When participants play multiple rounds of the ultimatum game with low stakes (the pot to be divided is $20), proposers offer much more than the minimum throughout all ten rounds of the experiment. However, when the stakes are high (e.g., the pot is $400) proposers offer less each round, producing a statistically significant convergence toward the theoretical prediction and responders continue to accept proposer's offers.

2.3.5 Statistical Issues

Consequential incentives can reduce noise and increase statistical power (Camerer & Hogarth, 1999; Smith & Walker, 1993). While consequential incentives cannot reduce the noise that enters data through unintended artifacts that stem from inadequate experimental design, they can reduce the noise that stems from under-motivated subjects, slow learning, hypothesis guessing, and in general those data artifacts that arise from subjects who don't put in their best effort in following experimenter instructions. This reduction in noise consequently leads to a reduction in Type II errors. The main benefit here is that when statistical power is increased, fewer subjects are needed to show significant relationships between dependent and independent variables.

2.3.6 Consequentialism in Consumer Research

Consequential incentives are more common in consumer research than abstraction and stationary replication and they seem to be gaining some traction there, but the vast majority of experiments in consumer research do not employ them, and of those that do, many employ incentives small enough to challenge the dominance criterion. This should be surprising. Many consumer decisions lead to different material consequences (i.e., the choice of changing a pension plan, betting $100 extra bucks on the roulette table, eating a chocolate Budino after dinner, etc.), which in turn may influence one's material and emotional well-being both in the short term and the long term. Such consequences imply that the use of consequential incentives would be highly compatible with consumer research conducted in the lab.

Furthermore, current consumer research presents a variety of commonly studied dependent variables that are compatible with the use of consequential incentives. Such dependent variables include preference measurements and related measures (e.g., preference, liking, choice, willingness to pay, purchase intent, attitudes, affect, etc.), evaluations (e.g., brand evaluations, product evaluations, attribute evaluations, advertisement evaluations, product quality evaluations, etc.), judgments (including forecasts or predictions), and tasks (e.g., memory and recall tasks, cognitive tasks, search tasks, clerical tasks, etc.). With respect to these dependent variables, BDM procedures can be used for preference and evaluation related measures, quadratic scoring for judgment measures (these two procedures will be addressed later in this article), and other forms of consequential incentives for tasks (discussed presently).

2.4 Types of Consequential Incentives

There are various types of tasks that are commonly employed in consumer research that are compatible with the use of consequential incentives (see Bonner, Hastie, Sprinkle, & Young (2000) for one taxonomy of incentive schemes). A short list of popular incentive schemes is provided here.

2.4.1 Piece-Rates

Piece-rates pay a predefined amount of money for each task, unit of output, or increased level of performance. For example, in a memory task, a participant might receive $0.25 for each successful digit they can recall in a memorized series of digits. Or in a market experiment, a participant might receive a $1.00 profit for every unit sold. This type of incentive scheme is perhaps the most basic type of incentive scheme that can be employed consequentially, and variants of this scheme are the most prevalent in experimental economics (Lee, 2007). Because this type of incentive scheme is very simple, it is easy to explain and easy to understand.

2.4.2 Thresholds

Goal- or quota-based incentives award a bonus for reaching some level of performance (i.e., the payment function has a discontinuous jump). Research suggests that thresholds also increase intrinsic motivation when compared to other incentive schemes (Bonner et al., 2000); however, because subject behavior may be significantly different when performance is close to the threshold, this payment scheme has limited use in the laboratory (Lee, 2007) unless it is specifically employed to study goals and quotas.

2.4.3 Probabilistic Incentives

Probabilistic incentive schemes employ payments that are paid with some probability between zero and one (i.e., lotteries). Such schemes can be used to reduce the costs of the researcher when an experiment involves large numbers of participants or multiple trials. For example, suppose ten participants must each make a real choice between two MP3 players. The cost of offering this choice to all participants can be significantly reduced by randomly choosing one participant per session to receive their chosen MP3 player. When an experiment involves multiple trials, costs can be reduced by choosing a subset of trials at random that will provide real payments. Uncertainty can also be used to incentivize competing goals. For example, Billeter, Kalra, and Loewenstein (2011) incentivize both prediction accuracy and performance in Snoddy's (1926) classic star tracing task. When subjects complete this task, they roll a die to see whether payment will be based on performance (a roll of 1–5) or both performance and accuracy (a roll of 6). This randomization scheme places heavy weight on performance, but also places some weight on prediction accuracy. Consequently, subjects are motivated to perform their best on both tasks (instead of manipulating their performance to match their accuracy predictions).

2.4.4 Competitive Incentives

Competitive incentive schemes pay participants based on their performance rank relative to other participants. For example, a memory task could be administered in a tournament with ten participants where the highest performer receives a cash prize of $100. A problem arises when payoffs are tournament-based. Participants who feel the likelihood of performing better than others (and thus receiving the prize) is very low may not be motivated to put in any effort (Bull, Schotter, & Weigelt, 1987; Dye, 1984; Friedman & Sunder, 1994). This lack of motivation would then violate the dominance criterion for these subjects. For this and other reasons (Dye, 1984), incentives based on individual performance are preferred in most cases to competitive incentives. The main exceptions to this guideline occur when a tournament incentive scheme is necessary to the experimental manipulation or when the experimental design employs tournaments in order to recreate real-world tournament-based incentives (Lim, Ahearne, & Sung, 2009; Lim, 2010).

2.4.5 Real Choices

Another possibility is to use real choices (e.g., a choice between two candy bars, a choice between multiple lotteries, a choice between possible courses of action, etc.). Importantly, when real choices are used, the difference in perceived value between the alternatives in the choice set should be large enough to meet the dominance criterion.

2.5 Absence of Deception

2.5.1 Explanation

The fourth important practice in an economic experiment is the absence of deception (Davis & Holt, 1993; Friedman & Sunder, 1994; Hey, 1998; Ledyard, 1995). Participants are never given a cover story, researchers never employ confederates to pose as participants, and while a researcher may not reveal all relevant information to participants during an experimental session, the information that is provided in either written or verbal form is

always true.

The "no deception" rule is particularly crucial when the instructions are related to *people* and *payoffs*. For instance, imagine a participant, who is told that she is playing a game with someone else, is in fact playing with a computer. Further imagine that this becomes a recurrent approach in the lab. Over time, suspicion might build causing participants to behave differently (e.g., cooperate more often in a prisoner's dilemma game when participants believe they are playing with a computer than when they believe the other player is a person (Andreoni & Miller, 1993)). Likewise, imagine a study participant is told the probability of winning a given gamble is 45%, but in fact it is 70%. Over time, participants may distrust information about payoff structures and again behave in a different manner. Since, participants often participate in multiple experiments conducted at the same lab, mistrust in the instructions is a real possibility and should be guarded against.

2.5.2 Absence of Deception in Consumer Research

Both the American Psychological Association and the American Sociological Association have rules of conduct stating that deception should only be used in experiments when the extent of the deception is justified by the expected benefits of the study and when effective non-deceptive alternative procedures are not feasible (APA, 2002; ASA, 1999). Despite these rules of conduct, the Journal of Experimental Social Psychology in 2002 published 117 studies, 53% of which used deception (Hertwig & Ortmann, 2008a). Epley and Huff (1998) analyzed the use of deception in the Journal of Personality and Social Psychology in 1996 and found that 42% of studies employed some form of deception. In marketing research the use of deception increased from 43% in 1975–1976 to 56% in 1996–1997 (Kimmel, 2001). The survey of the *Journal of Consumer Research* revealed that 38% of studies employ some form of deception, mostly in the form of cover stories. As Hertwig and Ortmann (2008b) state, "It is difficult to reconcile the still relatively high prevalence of (deception) with the notion that deception is reserved for those cases in which the study's prospective value justifies its use and effective alternatives are not feasible."

Although deception may be needed in some experiments in consumer research, it is possible to minimize its use under some circumstances without any detriment to the research. This is particularly true when consequential incentives are employed. For example, meaningful incentives can sometimes overcome the need for a cover story. Moreover, to minimize the impact and contamination of suspicion effects, an experimental lab can use separate participant pools that follow different criteria for the use of deceptive techniques.

3. Procedures to Elicit Preferences and Beliefs

In addition to the four practices just presented, consumer researchers often rely on measures of preference and beliefs to answer their research questions. Experimental economists often employ two procedures that can be used to elicit preferences (in the form of willingness to pay estimates) and beliefs (in the form of probability judgments) in a consequential manner.

3.1 The BDM Procedure

During an experiment where estimates of willingness to pay (WTP) for a good are elicited from participants, consequential incentives can be used to increase the likelihood that participants will provide honest, careful estimates. This is accomplished through the use of the BDM procedure (Becker, DeGroot, & Marschak, 1964). The procedure is clearly articulated in the following example script adapted from Wertenbrock and Skiera (2002):

> "How much money you are willing to spend for this (product). The purchase price is not yet determined. Once you state your highest willingness to pay, you may then draw a ball from this urn. The balls are labeled with different prices. If you draw a price that is less than or equal to the price you tell me, you will have to buy the (product) for the price you drew from the urn. If the price you draw is greater than the price you tell me, you will not be able to buy the (product). This procedure ensures that it is best for you to truthfully reveal the maximum price you are willing to pay. If you tell me a price that is higher, you may actually have to pay that higher price. If you tell me a price that is lower, you may be disappointed that you can't buy at a price that is lower than your "true" price. Note that you cannot influence the purchase price with the price you tell me. Because you draw the purchase price from the urn, it is completely random and independent of whatever you tell me. Do you have any questions? Now, what is the maximum amount you are willing to pay for this _____?" (p. 239–240)

The BDM procedure uses randomization to pit two opposing goals against each other (i.e., obtaining the good vs. paying a desirable price). Note that 1) it is important to ask for the "highest" price the participant is willing to pay, instead of simply for "how much" the participant is willing to pay; 2) it is important that the participant

recognizes that the procedure used to generate the actual purchase price is not influenced by her decision and that the price generated has a reasonable likelihood of being either above or below her stated price; and 3) it is critical to explain why the optimal strategy for any participant is to name her true willingness-to-pay. Of course, this script can be easily adapted to accommodate other randomization devices other than urns full of balls with prices on them.

This procedure can be easily applied to attitude or intention-like measures. For instance, consumer researchers often ask participants to indicate their liking for a good on a Likert scale. In addition to such questions which have no attached consequential incentives to be truthful, a research could also measure the participants' WTP using the BDM procedure.

3.2 Quadratic Scoring

When a meteorologist predicts a 20% chance of rain or a pundit predicts a 65% chance that a candidate will be elected, these are probability judgments. Such judgments can be useful in an experimental setting to elicit a participant's beliefs. In order to elicit truthful, effortful probability judgments from participants, a researcher can tie consequential incentives to a scoring rule.

A scoring rule assigns a score to a probability judgment based on its associated, observed outcome. For example, suppose an urn contains twenty red balls and ten white, and that the task is to judge the probability of drawing a red ball from the urn (without knowing beforehand the color distribution in the urn). One possible scoring rule for the probability judgment is to award a score equal to the participant's stated probability (p) if a red ball is drawn and to award a score of one minus the probability ($1 – p$) if a white ball is drawn. Such a scoring rule associates the probability judgment with the outcome of the random draw from the urn, but in this case, the rule will not likely induce the assessor to state her true belief about the likelihood of drawing a red ball. A participant who wishes to maximize her score under this rule is likely to predict extreme probabilities (e.g., $p = 1$, when her true belief is any probability between .5 and 1 or $p = 0$ when her true belief is any probability between .5 and 0 (Nelson & Bessler, 1989)).

In order to induce the participant to truthfully reveal her belief about the probability, a "strictly proper" scoring rule must be used. A scoring rule is "proper" if the assessor can only maximize her score for a probability distribution F by issuing a probabilistic forecast of F. In lay terms, a scoring rule is proper if it prompts forecasters to provide a forecast consistent with their beliefs. A rule is "strictly" proper if the maximum is unique (Gneiting & Raftery, 2007).

Though multiple strictly-proper scoring rules exist (e.g., quadratic, spherical, logarithmic, etc.; see Winkler et al. (1996) for a review of strictly proper scoring rules), the most commonly used proper scoring rule is the quadratic scoring rule (McKelvey & Page 1990; Nyarko & Schotter, 2002; Palfrey & Wang, 2009), where the payoff is determined by a quadratic function of the probability estimate. Table 1 is an example table that that implements a strictly-proper, quadratic scoring rule that assigns $\$(2p – p^2)$ if a blue ball is drawn from an urn of blue and red balls and assigns $\$(1 – p^2)$ if a red ball is drawn; p here represents the subject's predicted probability of drawing a blue ball. This table can be included with the following instructions to the respondent: "In the table below, circle the percentage in the first column that you think best represents the likelihood of drawing a blue ball from the urn. The amounts in the next two columns represent your payoffs in dollars if either a blue or a red ball is drawn."

In this example, if a participant predicts that the probability of drawing a blue ball is 100% and a blue ball is actually drawn, she receives $1.00. Alternatively, if a red ball is drawn, she receives nothing. Similarly, if she predicts the probability of drawing a blue ball is 30%, she receives $0.51 if a blue ball is drawn and $0.91 if a red ball is drawn.

Because this table implements a strictly proper scoring rule, it is always optimal for a subject to report their true belief. Also note that a researcher can provide the proper incentive without having to use the word "probability". The use of a scoring table removes this need as well as the need for a participant to perform any calculations. The task is relatively simple: choose from a menu of alternatives.

Table 1. Payoff schedule for different judgments using quadratic scoring

Likelihood of drawing Blue	Payoff if Blue is drawn	Payoff if Red is drawn
0%	0.00	1.00
10%	0.19	0.99
20%	0.36	0.96
30%	0.51	0.91
40%	0.64	0.84
50%	0.75	0.75
60%	0.84	0.64
70%	0.91	0.51
80%	0.96	0.36
90%	0.99	0.19
100%	1.00	0.00

4. Conclusion

Ariely and Norton state (2007), "The lack of communication between psychology and economics is particularly unfortunate because the fields share interest in similar topics that are of clear importance to public policy and social welfare; at the same time, however, the gaps in approach are substantial and epistemological, so bridging them is not trivial."

This article contributes to the important goal of bridging psychology and economics by presenting practices and procedures used by experimental economists to consumer researchers trained in the tradition of psychological experiments. The intent of this article is to prompt consumer researchers who have been trained in the norms of psychological experiments to allow these concepts to inform and augment their current methods and research. A coming-together of these two disciplines can significantly enhance our understanding of how consumers make decisions, how they interact with each other and with firms, and how changes in market rules and structure might influence consumer behavior. Furthermore, greater involvement of psychologically-trained consumer researchers in experimental economics will be important to understanding consumer behavior as well as improving the methodology of experimental economics.

References

American Psychological Association. (2002). Ethical Principles of Psychologists and Code of Conduct. Retrieved from http://www.apa.org/ethics/

American Sociological Association. (1999). Code of Ethics and Policies and Procedures of the ASA Committee on Professional Ethics. Retrieved from http://www.asanet.org/

Andrade, E. B., & Ho, T. H. (2007). How is the boss's mood today? I want a raise. *Psychological Science*, *18*(8), 668–671. http://dx.doi.org/10.1111/j.1467-9280.2007.01956.x

Andrade, E. B., & Ho, T. H. (2009). Gaming emotions in social interactions. *Journal of Consumer Research*, *36*(4), 539–552. http://dx.doi.org/10.1086/599221

Andrade, E. B., & Iyer, G. (2009). Planned versus actual betting in sequential gambles. *Journal of Marketing Research*, *46*, 372–383. http://dx.doi.org/10.1509/jmkr.46.3.372

Andreoni, J., & Miller, J. H. (1993). Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *The Economic Journal*, *103*(418), 570–585. http://dx.doi.org/10.2307/2234532

Ariely, D., & Norton, M. I. (2007). Psychology and Experimental Economics: A Gap in Abstraction. *Current Directions in Psychological Science*, *16*(6), 336–339. http://dx.doi.org/10.1111/j.1467-8721.2007.00531.x

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential

method. *Behavioral Science*, *9*(3), 226–232. http://dx.doi.org/10.1002/bs.3830090304

Billeter, D., Kalra, A., & Loewenstein, G. (2011). Underpredicting learning after initial experience with a product. *Journal of Consumer Research*, *37*(5), 723–736. http://dx.doi.org/10.1086/655862

Binmore, K. G. (1994). *Playing Fair*. Cambridge, MA: MIT Press.

Bonetti, S. (1998). Experimental economics and deception. *Journal of Economic Psychology*, *19*(3), 377–395. http://dx.doi.org/10.1016/S0167-4870(98)00012-9

Bonner, S. E., Hastie, R., Sprinkle, G. B., & Young, S. M. (2000). A review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. *Journal of Management Accounting Research*, *12*(1), 19–64. http://dx.doi.org/10.2308/jmar.2000.12.1.19

Bull, C., Schotter, A., & Weigelt, K. (1987). Tournaments and piece rates: An experimental study. *The Journal of Political Economy*, *95*(1), 1–33. http://dx.doi.org/10.1086/261439

Burnham, T., McCabe, K., & Smith, V. L. (2000). Friend-or-foe intentionality priming in an extensive form trust game. *Journal of Economic Behavior & Organization*, *43*(1), 57–73. http://dx.doi.org/10.1016/S0167-2681(00)00108-6

Camerer, C. (1997). *Rules for Experimenting in Psychology and Economics, and Why They Differ* (pp. 313–327). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-60495-9_25

Camerer, C. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

Camerer, C., & Ho, H. T. (1999). Experience-weighted Attraction Learning in Normal Form Games. *Econometrica*, *67*(4), 827–874. http://dx.doi.org/10.1111/1468-0262.00054

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*(1–3), 7–42. http://dx.doi.org/10.1023/A:1007850605129

Castellan, Jr, N. J. (1969). Effect of change of payoff in probability learning. *Journal of Experimental Psychology*, *79*, 178. http://dx.doi.org/10.1037/h0026962

Cohen, J. B., Pham, M. T., & Andrade, E. B. (2008). The nature and role of affect in consumer behavior. *Handbook of Consumer Psychology*, 297–348. Retrieved from http://www.columbia.edu/~tdp4/HCP2006.pdf

Davis, D. D., & Holt, C. A. (1993). *Experimental Economics*. Princeton, NJ: Princeton University Press.

Dye, R. A. (1984). The trouble with tournaments. *Economic Inquiry*, *22*(1), 147–149. http://dx.doi.org/10.1111/j.1465-7295.1984.tb00675.x

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*, 759–768. http://dx.doi.org/10.1177/0146167298247008

Fehr, E., & Zych, P. K. (1998). Do addicts behave rationally? *The Scandinavian Journal of Economics*, *100*(3), 643–661. http://dx.doi.org/10.1111/1467-9442.00127

Fehr, E., Gächter, S., & Kirchsteiger, G. (1997). Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica: Journal of the Econometric Society*, 833–860. http://dx.doi.org/10.2307/2171941

Friedman, D., & Sunder, S. (1994). *Experimental Methods: a Primer for Economists*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139174176

Güth, W., Ivanova-Stenzel, R., & Tjotta, S. (2004). Please marry me!—An experimental study of risking a joint venture. *Metroeconomica*, *55*(1), 1–21. http://dx.doi.org/10.1111/j.0026-1386.2004.00179.x

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378. http://dx.doi.org/10.1198/016214506000001437

Harrison, G. W. (1994). Expected utility theory and the experiments. *Empirical Economics*, *19*(2), 223–253. http://dx.doi.org/10.1007/BF01175873

Hertwig, R., & Ortmann, A. (2008a). Deception in experiments: Revisiting the arguments in its defense. *Ethics & Behavior*, *18*(1), 59–92. http://dx.doi.org/10.1080/10508420701712990

Hertwig, R., & Ortmann, A. (2008b). Deception in social psychological experiments: Two misconceptions and a research agenda. *Social Psychology Quarterly*, *71*(3), 222–227.

http://dx.doi.org/10.1177/019027250807100304

Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences*, *24*(3), 383–403. http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=88683

Hey, J. D. (1998). Experimental economics and deception: A comment. *Journal of Economic Psychology*, *19*(3), 397–401. http://dx.doi.org/10.1016/S0167-4870(98)00013-0

Holt, C. A. (1995). Industrial organization: A survey of laboratory research. In J. J. Kagel & A. E. Roth (Eds.), *The Handbook of Experimental Economics* (pp. 349–443). Princeton, NJ: Princeton University Press.

Holt, C., & Laury, S. (2002). Risk aversion and incentive effects. *The American Economic Review, 92*(5), 1644–1655. http://dx.doi.org/10.1257/000282802762024700

Irlenbusch, B., & Sutter, M. (2006). An experimental analysis of voting in the Stability and Growth Pact in the European Monetary Union. *Public Choice*, *129*(3/4), 417–434. http://dx.doi.org/10.1007/s11127-006-9064-6

Irwin, J., McClelland, G., & Schulze, W. (1992). Hypothetical and real consequences in experimental auctions for insurance against low-probability risks. *Journal of Behavioral Decision Making, 5*(2), 107–116. http://dx.doi.org/10.1002/bdm.3960050203

Jamal, K., & Sunder, S. (1991). Money vs gaming: Effects of salient monetary payments in double oral auctions*. *Organizational Behavior and Human Decision Processes, 49*(1), 151–166. http://dx.doi.org/10.1016/0749-5978(91)90046-V

Kachelmeier, S., & Shehata, M. (1997). Internal auditing and voluntary cooperation in firms: A cross-cultural experiment. *The Accounting Review, 72*(3), 407–431. http://www.jstor.org/stable/248478

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*(4), 237–251. http://dx.doi.org/10.1037/h0034747

Kimmel, A. (2001). Deception in marketing research and practice: An introduction. *Psychology and Marketing, 18*(7), 657–661. http://dx.doi.org/10.1002/mar.1024

Koehler, J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences, 19*(1), 1–16. http://dx.doi.org/10.1017/S0140525X00041157

Kruglanski, A. (1975). The human subject in the psychology experiment: fact and fiction. *Advances in Experimental Social Psychology, 8*. http://dx.doi.org/10.1016/S0065-2601(08)60249-X

Ledyard, J. (1995). *Public Goods: A Survey of Experimental Research.* Princeton University Press.

Lee, J. (2007). Repetition and Financial Incentives in Economics Experiments. *Journal of Economic Surveys, 21*(3), 628–681. http://dx.doi.org/10.1111/j.1467-6419.2007.00516.x

Liberman, V., Samuels, S., & Ross, L. (2004). The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. *Personality and Social Psychology Bulletin, 30*(9), 1175. http://dx.doi.org/10.1177/0146167204264004

Lim, N. (2010). Social loss aversion and optimal contest design. *Journal of Marketing Research*, *47*(4), 777–787. http://dx.doi.org/10.1509/jmkr.47.4.777

Lim, N., Ahearne, M. J., & Ham, S. H. (2009). Designing sales contests: Does the prize structure matter? *Journal of Marketing Research*, *46*(3), 356–371. http://dx.doi.org/10.1509/jmkr.46.3.356

List, J., & Cherry, T. (2000). Learning to accept in ultimatum games: Evidence from an experimental design that generates low offers. *Experimental Economics, 3*(1), 11–29. http://dx.doi.org/10.1007/BF01669205

Loewenstein, G., & Adler, D. (1995). A bias in the prediction of tastes. *The Economic Journal, 105*(431), 929–937. http://dx.doi.org/10.2307/2235159

Loewenstein, G., & Schkade, D. (1999). Wouldn't it be nice? Predicting future feelings. *Well-being: The Foundations of Hedonic Psychology*, 85–105.

McKelvey, R., & Page, T. (1990). Public and private information: An experimental study of information pooling. *Econometrica , 58*(6), 1321–1339. http://dx.doi.org/10.2307/2938318

Nelson, R., & Bessler, D. (1989). Subjective probabilities and scoring rules: Experimental evidence. *American Journal of Agricultural Economics, 71*(2), 363–369. http://dx.doi.org/10.2307/1241594

Nyarko, Y., & Schotter, A. (2002). An experimental study of belief learning using elicited beliefs. *Econometrica,*

971–1005. http://dx.doi.org/10.1111/1468-0262.00316

Palfrey, T. R., & Wang, S. W. (2009). On eliciting beliefs in strategic games. *Journal of Economic Behavior & Organization*, *71*(2), 98–109. http://dx.doi.org/10.1016/j.jebo.2009.03.025

Phillips, L., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology, 72*(3), 346–354. http://dx.doi.org/10.1037/h0023653

Potters, J., & Van Winden, F. (2000). Professionals and students in a lobbying experiment Professional rules of conduct and subject surrogacy. *Journal of Economic Behavior and Organization, 43*(4), 499–522. http://dx.doi.org/10.1016/S0167-2681(00)00133-5

Ross, L., & Ward, A. (1996). Naive realism in everyday life: Implications for social conflict and misunderstanding. *Values and Knowledge,* 103–135.

Sawyer, A. (1975). Demand artifacts in laboratory experiments in consumer research. *Journal of Consumer Research,* 20–30. http://dx.doi.org/10.1086/208604

Shimp, T., Hyatt, E., & Snyder, D. (1991). A critical appraisal of demand artifacts in consumer research. *Journal of Consumer Research,* 273–283. http://dx.doi.org/10.1086/209259

Smith, V. (1982). Microeconomic Systems as an Experimental Science. *The American Economic Review, 72*(5), 923–955. Retrieved from http://www.jstor.org/stable/1812014

Smith, V., & Walker, J. (1993). Monetary Rewards and Decision Cost in Experimental Economics. *Economic Inquiry, 31*(2), 245–261. http://dx.doi.org/10.1111/j.1465-7295.1993.tb00881.x

Snoddy, G. (1926). Learning and stability: A psychophysiological analysis of a case of motor learning with clinical applications. *Journal of Applied Psychology, 10*(1), 1–36. http://dx.doi.org/10.1037/h0075814

Starmer, C. (1999). Experiments in economics: should we trust the dismal scientists in white coats? *Journal of Economic Methodology, 6*(1), 1–30. http://dx.doi.org/10.1080/13501789900000001

Stricker, L., Messick, S., & Jackson, D. (1969). Evaluating deception in psychological research. *Psychological Bulletin, 71*(5), 343–351. http://dx.doi.org/10.1037/h0027232

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453. http://dx.doi.org/10.1126/science.7455683

Van Boven, L., & Kane, J. (2006). Predicting feelings versus choices. *Judgments over Time: The Interplay of Thoughts, Feelings, and Behaviors,* 67–81.

Wertenbroch, K., & Skiera, B. (2002). Measuring consumers' willingness to pay at the point of purchase. *Journal of Marketing Research, 39*(2), 228–241. http://dx.doi.org/10.1509/jmkr.39.2.228.19086

Winkler, R., Muñoz, J., Cervera, J., Bernardo, J., Blattenberger, G., Kadane, J., & Ríos-Insua, D. (1996). Scoring rules and the evaluation of probabilities. *Test, 5*(1), 1–60. http://dx.doi.org/10.1007/BF02562681

Copyrights