Social Desirability Bias in Willingness-to-Pay for Products with Normative Attributes

Jayson L. Lusk
Professor and Willard Sparks Endowed Chair of Agribusiness
Department of Agricultural Economics
Oklahoma State University

F. Bailey Norwood
Assistant Professor
Department of Agricultural Economics
Oklahoma State University


Copyright 2006 by Jayson L. Lusk and F. Bailey Norwood. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright appears on all such copies.
Social Desirability Bias in Willingness-to-Pay for Products with Normative Attributes

Assessing the marketability of new agricultural products is of interest to both agribusinesses and agricultural economists. Agribusinesses’ motivation is straightforward: successful new products translates into greater profits. Agricultural economists (hereafter, economists) share this motivation at times, acting as a consultant to firms. However, is more typical for economists to conduct marketing studies with a policy implication in mind.

Consider the following examples. The United States Department of Agriculture grading system is intended to provide consumers with the meat quality information they need for purchasing decisions. To better identify consumer preferences for meat, economists study consumers’ willingness-to-pay (WTP) for guaranteed tender steaks and steak flavor (Lusk et al.; Lusk, Feldkamp, and Schroeder; Feuz et al.). Government regulation of genetically modified food should be formed based on consumer attitudes towards such food. Thus the over 25 marketing studies measuring consumer attitudes towards genetically modified food (for a summary of these studies see Lusk et al. 2005). Food regulations have the potential to create food value and raise food costs. Good regulations, most economists would say, are those that create more value than costs. Many economists spend their career measuring these values. Economists have measured the additional value (or lack thereof) in food created by Country-of-Origin Labeling (Tonsor et al.; Loureiro and Umbarger), bans on antibiotic use in swine production (Lusk, Norwood, and Pruitt), animal tracking systems (Dickinson and Bailey 2002 and 2005) and environmentally friendly foods (Loureiro et al.).

Whether an economist or agribusiness is conducting a marketing study, the study’s objectives usually include (1) product design and/or (2) evaluate the potential success of a particular product design. This is true for the firm concerned whether the product can generate profits or the economist concerned with whether consumers are willing to pay the additional costs imposed by a new governmental regulation. Put differently, the marketing study is a prediction exercise; the goal is to predict the success of a product in a real consumer setting.

The ability of a marketing study to predict behavior in a real consumer setting is referred to as external validity. For example, if 30% of subjects in a marketing study prefer an antibiotic-free pork chop to a regular pork chop at a $1.00 per pound premium, one would hope that antibiotic-free pork sold in grocery stores at a $1.00 per pound premium would accrue a 30% market share in pork chop sales. Of all the marketing research methods from survey questionnaires, laboratory experiments, and field experiments, the method exhibiting the greatest external validity is a test market. Because test markets place an item on sale in a real consumer setting, it provides accurate predictions of sales in other real consumer settings. Some might argue that a
test market is the “real thing”, and by definition is 100% accurate. When predicting real behavior, it is always best to use observations of real data. To predict the marketability of a food product, it is difficult to beat data on actual sales in actual grocery store.

In economics lingo, a test market is a natural field experiment (Harrison and List) in that the subjects are in a natural setting and are unaware they are being observed. While natural field experiments have strong external validity, other methods with weaker external validity are often employed. In fact, survey questionnaires, laboratory experiments, and framed field experiments constitute the majority of agricultural economic studies. The main reason is cost. Economists like to measure consumers’ response to different prices, yet few have the money to place items on sale at different prices at different locations. It is far cheaper to recruit a pool of subjects and expose them to different prices. There are other drawbacks to test marketing. Competitors can see a test market yet they cannot see a laboratory experiment, and firms often conduct market research before the product is even able to be produced, which negates the possibility of test marketing.

**Social Desirability Bias in Marketing Research**

The most common marketing research tools in agricultural economics are survey questionnaires, laboratory experiments, and framed field experiments. Examples abound: Keneko and Chern used a telephone survey asking respondents if they would pay less for non-genetically modified (GM) salmon; Dickenson and Bailey used auctions in a laboratory experiment to measure the value consumers place on traceable meat; and Rousu et al. utilized a framed field experiment where shoppers in a grocery store submitted bids on GM cigarettes. The purpose of these studies is to assess how consumers would respond to GM salmon, traceable meat, and GM cigarettes should they be offered in the marketplace. Each study provides useful information, but share a common potential bias. In all three methods the subjects know they are participating in an experiment or survey, and subjects behave differently when they know they are being watched (Levitt and List).

When the good being considered has normative implications (i.e. generates positive or negative externalities), subjects in experiments or surveys may exhibit a social desirability bias (SDB), where they try to create a favorable impression on the research administrator. Subjects who believe GM food creates dangers for others may want to give the impression that they oppose GM food even if they have no stance on the issue. Similarly, if traceable meat is presented as safer, people may be reluctant to say they do not care about food safety. The SDB bias—to our knowledge—was first addressed by psychologists and sociologists. One example is the study of racial attitudes. One can ask people about their attitudes towards other races, but people are likely to hide any racist tendencies they may possess (Plant, Devine, and Brazy).
In 1960 two psychologists developed a scale to measure SDB (Crowne and Marlowe). Referred to as the Marlowe-Crowne Social Desirability Bias (MCSDB) scale, it garnered the attention of marketing specialists curious as to whether SDB tainted marketing studies. Indeed, half of the marketing studies using SDB scales found that behavior in marketing studies was correlated with the SDB measures (King and Bruner), leading some to conclude that “social-desirability bias is considered to be one of the most common and pervasive sources of bias affecting the validity of experimental and survey research findings in psychology and the social sciences,” (King and Bruner, page 80).

Consider two examples in the economics literature. A contingent valuation study by Leggett et al. found that people expressed higher values for visits to Fort Sumter, South Carolina when asked in face-to-face interviews than when the survey was self-administered, despite the fact that in neither case the interviewer knew the subjects’ answers. Lusk, Norwood, and Pruitt conducted a framed field experiment where consumers were recruited and administered a choice experiment. Subjects chose between a regular pork chop plus money or an antibiotic-free (AF) pork chop, eliciting their willingness-to-pay for the AF pork chop. AF pork is thought to reduce health hazards to society. Subjects had to inform the administrator which pork chop they chose, and subjects surely knew that the researchers were interested in the marketability of AF pork. It is not far-fetched to assume they were more likely to chose the AF pork simply to please the researchers. Later, AF pork was placed on sale along with regular pork chops, so consumers could purchase or not purchase AF pork without being watched. The proportion of pork chop sales devoted towards AF pork was generally lower than what the choice experiment would predict. Another exhibit of SDB, the authors found that “subjects exhibit stronger moralist and pro-social behavior when they know they are being scrutinized,” (Lusk, Pruitt, and Norwood).

The presence of SDB limits external validity of marketing research methods other than natural field experiments. Yet, economists have done little to mitigate its effect. The purpose of this article is to review how other sciences have dealt with SDB, and propose a few research design that may dampen its impact.

**Dealing with Social Desirability Bias in Marketing Research**

As mentioned previously, the first marketing studies to address social desirability bias (SDB) did so by evaluating how subjects’ answer to questions differed as their social desirability level as measured by the Marlowe-Crowne SDB scale varied. This scale asks subjects to read 33 statements of personal attitudes and to answer true or false as to whether it describes the subjects’ personality. Each item contains an attitude or activity that is thought to “meet the criterion of cultural approval and yet be untrue of virtually all people” (Crowne and Marlowe, 1964, page 22). For example, the fourth item states,
“I have never intensely disliked anyone.” To answer true would meet cultural approval, but hardly no one can honestly answer true to this question. Thus, a person who answers “true” is likely exhibiting social desirability bias. By adding up the number of times an individuals’ answer is socially desirable but unlikely to be true, one has a measure of that person’s social desirability bias (Crowne and Marlowe 1960; 1964).

Assuming the MCSDB scale (other, similar scales are also available) does indeed measure SDB, it can be used to as an explanatory variable much like demographic variables. For example, suppose one collects auction bids in a laboratory experiment for a good with positive externalities, such as environmentally friendly dishwashing liquid. Along with these bids the MCSDB scale is collected for each individual, allowing one to estimate the following regression.

\[(1) \quad Bid_i = \alpha_0 + \alpha_1(MCSDB_i) + \epsilon_i\]

where \(Bid_i\) is the \(i^{th}\) individual’s bid and \(MCSDB_i\) is their score on the MCSDB scale which ranges from zero to 33. The average bid can be made conditional on the degree of social desirability exhibited, but if one is using the bids to assess the marketability of the product in a real consumer setting, it is unclear to what value \(MCSDB_i\) should be set. As a rough estimate, one could evaluate \(MCSDB_i\) two standard deviations below and above the sample mean to obtain a prediction interval. In the psychology literature some have simply set \(MCSDB_i\) to zero (Reynolds) to remove the effects of SDB on a variable. This use of scales to correct for a bias is similar to the certainty-calibrations used to correct for hypothetical bias in economic studies (e.g. Champ and Bishop; Johannesson et al.; Norwood Luter and Massey; aNorwood et al.) except external validity of the latter has received more attention.

Perhaps a more rigorous approach to removing SDB is to develop an understanding of the motivations that drive consumer behavior. Miniard and Cohen formalize a consumer motivation model that distinguishes between personal and normative motivations. To paraphrase their model in economics lingo, they assume that goods provide two types of utility. One is the familiar direct utility; the personal happiness individuals receive from consumption, regardless of social influences. What we call direct utility Miniard and Cohen call personal motivation. They also assume that individuals receive utility from being viewed favorably by society. While few economists dispute that such utility may exist, they rarely incorporate it into economic models. Fisher summarizes what Miniard and Cohen call normative motivations as, “...This component reflects the extent to which individuals are motivated by the expectations of another person or group,” (Fisher, page 305). Fisher argues that SDB should influence normative motivation but not personal motivation. It seems by definition alone that Fisher is correct.
Fisher then tests the extent to which indirect questioning reduces SDB by examining the degree to which it influences questions with normative motivation versus questions with only personal motivation. For example, students were presented with cordless headphones (a “new” product that did not really exist) and administered questions regarding their attitude towards the product. Some students were directly asked to indicate their willingness to purchase the product if it becomes available and others were asked indirectly if they thought the typical student would purchase it. They were then asked to indicate the extent to which they agreed with a series of statements. Some statements reflected personal motivation, such as “the new headphones provide more freedom of movement than other headphones.” Other statements reflected normative motivations, such as “students I know would have a favorable reaction if I bought one of the new products.” Again, some students were asked their personal opinion, and others were asked to answer as they think the typical college student would.

The study revealed that indirect questioning reduced the extent to which people agreed with the normative statements. Fisher interprets this to imply that indirect questioning reduces social desirability bias. We believe these findings suggest an alternative format for eliciting people’s willingness-to-pay. Instead of asking how much individual’s themselves value a good, ask them how much the “typical person” or some other group values the good. Since the individual is now assessing how others behave, one cannot modify their answers to create a favorable impression of themselves, and social desirability bias may be removed; or at least, partially mitigated.

Consider two different methods for predicting the outcome of a political election. One is polling individuals; simply asking them who they would vote for. Evidence suggests extensive SDB in such questions, as studies have shown people will say they voted in a recent election when in fact they did not, just to “look good” to an interviewer (Bertrand and Mullainathan). It is therefore likely that many people answer polls saying they will vote for someone, but in fact do not vote at all. If the probability of being one of these people is correlated with the probability of voting for a particular candidate (given that one does vote) then polls are biased.

A second method uses prediction markets to assess the outcome of an election. The Iowa Electronic Markets are one well-known example. Instead of asking people who they will vote for, the researcher asks people to predict who other people will vote for, and pays them based on the accuracy of their predictions. This is akin to indirect questioning, and studies have shown that such markets outperform polls in predicting election outcomes (Berg et al.). Indeed, research has revealed that the aggregate belief of a non-small and diverse group can provide surprisingly accurate predictions of outcomes. For example, prediction markets have been constructed to forecast the outcome of football games; however, some research suggests that a simple average of experts’ opinions forecasts equally well (Chen et al.). In fact, simple methods of
aggregating individuals’ opinions have a long history of providing good forecasts of events. British scientist Francis Galton documented one such incident in 1906. A contest was held at a country fair where individuals guessed the weight of an ox and prizes were given for the best guesses. A total of 800 people participated. Although few of the participants were experienced butchers, the average guess of 1,198 lbs was shockingly close to the true weight of 1,197 lbs (Surowiecki).

A more interesting and remarkable example comes from the disappearance of the U.S. submarine Scorpion in 1968. The Navy only knew the sunken submarines’ last reported location, which narrowed the search radius to 20 square miles. Naval Officer John Craven compiled a diverse group of people including mathematicians, submarine specialists, and salvage men and asked the group to place wagers on the cause of the submarine’s difficulty and the speed it sank, among other items. Bayes’ Theorem was used to aggregate the wagers to predict a single location. Although no single individual identified the location, the submarine was found only 220 yards from the aggregate, composite prediction (Surowiecki).

If the aggregate belief predicts outcomes well, and if predicting other peoples’ behavior results in less SDB than predicting one’s own behavior, marketing studies may be improved by asking subjects how other people would behave, instead of the person’s own attitude towards a new product. This is especially true for goods with normative attributes, like those associated with an externality. We refer to values elicited when you ask people to predict other peoples’ value, as opposed to observing their own behavior where the context encourages SDB, as indirect valuation. Whether indirect valuation improves marketing studies of new products has yet to be tested. The purpose of this paper is to motivate research testing such hypotheses. Below, several methods for conducting indirect valuation are discussed.

**Examples of Indirect Valuation**

Numerous methods for conducting valuation studies exist. Here we concentrate on those most commonly conducted in laboratory or framed field experiments. It is assumed throughout that the goal of the research is to predict consumer behavior in an actual consumer setting via laboratory or framed field experiments. Consider choice experiments where consumers choose from two or more products with varying attributes and different prices. Subjects make their choices and the choice data are used to estimate a random utility function for products and product attributes. As indicated earlier, unless subjects make choices without knowing they are being watched, their choices may reflect a social desirability bias (SDB). Specifically, they may be more willing to pay for a good with positive externalities in the experiment than they are in a real shopping experiment where they are not being watched.
The indirect alternative to choice experiments is to present the same product choices at the same prices and ask which product the subject believes will elicit greater sales, rather than which product the consumer wishes to purchase. Subjects will now view the purchasing decision outside of the context of an observed experiment, and outside of themselves where SDB exists. Answers to this experiment can be used to construct a random utility function just like the normal choice experiment.

Auctions are another popular valuation method. Subjects are presented a good and are asked to submit bids to an incentive-compatible auction like the BDM or nth price auction. The indirect counterpart to auctions could consist of asking subjects to submit a price they believe half of consumers would purchase and half would not. The bids submitted are the subjects’ opinion of other consumers’ mean willingness-to-pay (WTP). Going a step further, one could ask subjects to submit bids they believe would induce X% to purchase the item. Combining these bids with the previous bids and assuming a two parameter distribution for WTP allows one to calculate the predicted WTP distribution.

Often economists prefer individuals to experience a monetary outcome for their actions to encourage them to take the question seriously. Ideally, subjects would be asked to predict consumer behavior and would be rewarded based on their prediction accuracy. However, this would require a test market to accompany the valuation experiment, and as discussed earlier, if a test market is to be conducted there may be little need for a laboratory or framed field experiment. Test markets are the ideal “experiment”, and other experiments are usually conducted only as a second-best alternative.

An alternative is to reward subjects' predictions based on their conformity with “expert predictions.” The subjects are the real expert in valuation experiments, but they do not necessarily know this. If subject A truly believes product B will have a 60% market share at a particular price, if asked her predicted market share and told she will be compensated based on her prediction accuracy, it is unlikely her answer would change if accuracy is based on a true test market outcome or the expert prediction. In the latter case, the expert prediction is a farce, but a farce that may produce accurate valuation assessments.

In the choice experiment, the individual chooses which item she believes will have the greatest sales. Compensation can be made by giving her a sum of money for each “true” answer. In the auction, individuals can be compensated based on the difference between their bid and the “true” price that induces half of consumers to make the purchase. In both cases, the “true” answer or price may be substituted with expert opinion.

Prediction markets may also appear to be a useful information elicitation format. One could have individuals buy and trade contracts whose value is unknown at the time of trading, but whose future value will be tied to sales of a new product. For
example, one could equate the contract’s value equal to the market share of the product in a test market, or expert opinion of what that market share would be (again, substituting the “truth” with a truth-proxy). However, some of our preliminary research suggests that prediction markets are inferior to simply asking people their predicted market share. Our experiments allowed students to guess the market share of antibiotic-free pork chops at a local grocery store relative to total pork chop sales. The antibiotic-free pork chops were test-marketed over a two-week period. Then, they traded contracts whose future value equaled the true market share of the antibiotic-free pork chops. The price of a contract then constituted a prediction of the market share. For example, if the contracts trade at $0.30, the market believes the market share will be 30%. Their simple guesses were generally more accurate that market prices. Based on our experience conducting the experiment, the reason is that students who have a difficult time understanding the prediction market tend to participate at the end of the experiment, simply because they want to participate and they feel time is running out, and they will trade at virtually any price. Their enthusiasm to trade overwhelms their enthusiasm for making money, leading to little information reflected in the market price.

**Synergies of Indirect Valuation**

The ideas put forth above are just that: ideas. No study has yet determined whether indirect valuation will produce numbers that better match consumer behavior in a real shopping setting. However, in our opinion the ideas are strong enough that such studies are warranted. Indirect valuation has been considered, though only in hypothetical settings. In a recent experiment we considered the use of indirect valuation to correct for hypothetical bias (Lusk and Norwood). Hypothetical bias is the phenomena where individuals’ state they will pay a higher price for an item than their actual behavior reveals. We conducted an experiment where subjects were given $10 for participating in an experiment. A vote was taken on whether each student would give up $3.00 in exchange for a plant. If more than 50% voted yes, each student gave up the $3.00 and received a plant regardless of how they voted individually. Subjects were told the plants would be destroyed if the proposition failed. A total of 37% of students voted yes, so the proposition failed.

Another similar group of subjects were asked to vote on the same proposition, except that it was hypothetical. Although they were asked to vote as if money would really be exchanged for the plants, no exchange would really take place. In this case 67% of subjects voted yes, a clear indication of hypothetical bias. Finally, the third and fourth groups of similar subjects were told about the real proposition in the first group, but were not told the outcome. Instead, the third group was asked what percent of subjects they believed voted yes. They predicted 37%—exactly equal to the true percent!
The fourth group was asked to predict the “expert opinion” of the percent voting yes, and they predicted 33%, which is not statistically different from the true percentage.

Thus, the limited evidence thus far indicates indirect valuation can correct for hypothetical bias. If indirect valuation could also correct for social desirability bias it would become even more valuable as a research tool, as it may accurately assess the marketability of products by observing subject behavior in hypothetical situations and without the effect of social desirability bias.

Summary Implications

When predicting consumer shopping behavior from experiments, economists often insist that the experiment subjects make real choices where real products are exchange for the subjects’ money. The literature clearly demonstrates that the absence of real money creates a hypothetical bias, and a misleading picture of true consumer behavior. The purpose of this paper is to argue that most experiments are still subject to another form of bias: social desirability bias. Only natural field experiments are free from this bias, the reason being that in natural field experiments the subjects are unaware they are even in an experiment. They do not know they are being watched by the research, and are unaware their behavior is recorded.

Most marketing studies using economic experiments are classified as laboratory, artefactual, or framed field experiments. In all cases the subjects are aware they are being watched, and may tend to behave in ways that they believe create a favorable impression. This is especially true for goods associated with positive or negative externalities, where behavior has normative or social implications. Social desirability bias has received much attention in the marketing literature but little in economics where non-market valuation studies of normative goods abounds.

This paper discusses two methods for alleviating social desirability bias, both used extensively in the marketing literature. One is to control for individuals’ social desirability level using the Marlowe-Crowne social desirability scale as a control variable. The other is to ask indirect questions instead of direct questions. In non-market valuation, this amounts to asking individuals how they perceive others value a good, instead of how the individual’s personal value. Combining this with the fact that people often predict others’ behavior better than their own, and it may remedy hypothetical bias, this indirect valuation method may provide improved predictions of a new product’s sales.

It should be noted that a number of smaller design issues can also dampen social desirability bias. Studies have documented that preserving individuals’ anonymity and self-administered (as opposed to interviewer-administered) surveys reduce the bias. Every study should have a script informing the subject to avoid social desirability bias by telling them the research is not trying to persuade them to do or believe anything
other than express their true, personal preferences. This paper is particularly concerned with the potential use of indirect valuation though, and we have plans for experiments to evaluate its potential. We encourage other researchers to do the same, and to search for alternatives to our suggestions.
References


