

# Interviewer Effects in an Environmental Valuation Telephone Survey

Min Gong

Center for Research on Environmental Decisions, Columbia University

406 Schermerhorn Hall - MC 5501

1190 Amsterdam Ave

New York, NY 10027

David Aadland\*

Department of Economics and Finance, University of Wyoming

1000 E. University Avenue

Laramie, WY 82071

August 20, 2009

**Abstract.** We investigate interviewer effects in a random digit dial telephone survey of recycling valuation and behavior. Lack of market experience with environmental goods may cause respondents to be more prone to interviewer effects during valuation surveys. Following previous research in both psychology and survey methodology, we test the direct effects of interviewer gender and race, as well as the interaction effects between interviewer and respondent characteristics. Using data from 132 interviewers and 3,870 interviewees, we apply a hierarchical regression model that accounts for the clustering of interviews and controls for a variety of other confounding variables. We confirm the existence of both direct and conditional interviewer effects. Respondents state higher willingness to pay when interviewed by white or female interviewers. There were also significant interaction effects between interviewer and respondent characteristics. The directions of the interviewer effects are consistent with previous survey research and social psychology theories. We also find “timing effects” related to the time and date of interviews that are not traditionally included in studies of interviewer effects.

**Keywords:** interviewer effects, willingness to pay, recycling, social attribution model

---

\*The authors are post-doctoral fellow at the Center for Research on Environmental Decisions, Columbia University, and associate professor in the Department of Economics and Finance, University of Wyoming, respectively. Contact: [aadland@uwyo.edu](mailto:aadland@uwyo.edu). Phone: 307-7664931. This research is supported by National Science Foundation grant #0108159.

## 1. Introduction

Interviewer effects are a common problem in survey work. Respondents, especially during interviews with socially sensitive content or unfamiliar topics, are often influenced by the interview process and characteristics of the interviewer. Previous research has shown that respondents may be influenced by the gender of the interviewer (Landis *et al.*, 1973; Groves and Fultz, 1985), age of the interviewer (Norris and Hatcher, 1994), race of the interviewer (Cotter *et al.*, 1982), interviewer's information and expectation (Singer *et al.*, 1983), interview style (Dijkstra, 1987), interview language and intonation (Barath and Cannell, 1976; Lee, 2001) and social distance between the interviewer and respondent (Williams, 1964,1968).

There are two common approaches in the literature for analyzing interviewer effects. In the first approach, interviewer characteristics have a direct effect on individual responses in the sense that "respondents evaluate interviewers and make inferences about their norms, values, and beliefs based upon their observable characteristics" and may "tailor their answers in an effort to elicit interviewer approval by conforming to their perceived expectations" (Fendrich *et al.*, 1999). In other words, there exists consistent social desirability bias that can be identified and predicted by studying interviewers' characteristics. Fendrich *et al.* label this approach the *Social Attribution Model*.

According to the *Social Attribution Model*, interviewer race (gender) effects are especially important in surveys related to race (gender). This prediction has been confirmed empirically by several studies. For example, Finkel *et al.* (1991) show in a telephone pre-election survey that White respondents were 8-11% more likely to voice support for an African American candidate when interviewed by an African American interviewer than by a White interviewer. Similarly, respondents were also more likely to voice support for a White candidate when interviewed by a White interviewer than by an African American interviewer.

Landis *et al.* (1973) find that female respondents reported more feminist responses on liberation items when interviewed by a male interviewer than by a female interviewer.

Although the *Social Attribution Model* focuses on race and gender issues, the model also applies to other topics provided respondents can infer the interviewer's attitude on the topic from perceived interviewer characteristics. For example, Groves and Fultz (1985) found that respondents stated a more optimistic economic outlook when interviewed by a male interviewer than a female interviewer. The authors argued that respondents believed that male interviewers were generally more optimistic about economic conditions and as a result were unconsciously influenced by male interviewers' optimism.

The second approach for characterizing interviewer effects proposes that "judgments about interviewers vary according to subject characteristics" (Fendrich *et al.*, 1999). That is, there is an interaction between interviewers and respondents that can be associated with certain characteristics of the respondents. Fendrich *et al.* label this approach the *Conditional Social Attribution Model*. Previous research often relies on social distance theory (Williams 1964) to explain these interactions. The greater the social distance between respondent and interviewer, the more likely a respondent alters her answer to conform to the interviewer. Social distance is usually measured in terms of differences or similarities in the social identities of interviewers and respondents (Johnson *et al.*, 2000). For instance, the social distance of an interviewer and a respondent with different genders is greater than when an interviewer and a respondent are of the same gender.

Another possible explanation for these interactions is that respondents may have different propensities to conform and therefore differ in their susceptibility to interviewer effects. For example, in a review study on propensity to conform across genders, Eagly (1978) reports that 23 previous studies show females are more likely to conform than men. Thus, female respondents may be more susceptible to interviewer effects than male

respondents. As we will discuss later, the predictions from social distance theory and differences in the propensity to conform may contradict one another, allowing us to test two competing hypotheses.

Interviewer effects have also been studied over a broad range of topics, including market research (Udow, 1942; McKenzie, 1977), health surveys (Clarke, 2003), drug use (Johnson *et al.*, 2000), and political opinions (Finkel *et al.*, 1991), to mention a few. Recently, there has become interest in interviewer effects in valuation surveys for non-market goods (Bateman and Mawby, 2004; Loureiro and Lotade, 2005; Loomis *et al.*, 2006).<sup>1</sup> Unlike the valuation of private goods, where agents experience repeated market interactions with the good, respondents in environmental valuation surveys often have little-to-no experience with the good. As a result, respondents are likely to be uncertain about their subjective value for the environmental good and be tempted to rely on subtle cues from the interviewer or preconceived notions about what the interviewer views as a “good” answer. This process, whereby respondents update their subjective values for an environmental good – either consciously or subconsciously – in response to the manner in which the interview or valuation questions are framed can be thought of as a form of updating beliefs (McFadden, 2001; Aadland *et al.*, 2007).

To the extent that interviewer effects in environmental valuation surveys lead to biased estimates of the benefits of environmental programs, they have the potential to create significant distortions in the allocation of public resources. In their panel report to the NOAA in 1993, Arrow *et al.* (1993) acknowledge the possibility and importance of interviewer effects in contingent valuations for environmental goods, suggesting that “major

---

<sup>1</sup> Among all environmental valuation methods, the use of surveys is probably the most common method to assess natural-resource damages and to guide public policy by federal and state agencies. The best-known example is the damage assessment for the 1989 Exxon Valdez oil spill (Carson *et al.*, 1992). Other prominent valuation surveys involve air pollution in Southern California (Brookshire *et al.*, 1982), national water quality benefits from the Clean Water Act (Carson and Mitchell, 1993) and mitigation of climate change (Hersch and Viscusi, 2006).

CV [contingent valuation] studies should incorporate experiments that assess interviewer effects.”

Although few economic studies are devoted exclusively to interviewer effects, several systematic interviewer effects have been identified in previous valuation surveys. For example, both Desvousges *et al.* (1998) and Boyle and Bishop (1988) observe the existence of individual difference among interviewers. Walsh *et al.* (1990) report that willingness to pay (WTP) elicited by an interviewer with no economic experience is lower than that elicited from three interviewers who had training in economics. Bateman and Mawby (2004) illustrate that different attire styles on the same interviewer elicit different WTP for an environmental good. Loomis *et al.* (2006) find that language influences WTP responses in a telephone survey of fire protection policy. Surprisingly, there has been little research on gender or race interviewer effects in valuation surveys. These are the two characteristics most commonly studied in the general interviewer-effects literature.

To the best of our knowledge, the only paper that focuses on gender or race interviewer effects is Loureiro and Lotade (2005). The authors compare the WTP estimates of fair trade, shaded and organic coffee labels from two interviewers in face-to-face interviews. Both interviewers are males, one white from the United States and one black originally from Africa. No other characteristics of the interviewers are given. The two interviewers follow a transcript during the interview process. Loureiro and Lotade find a significant difference between the WTP elicited by the two interviewers, especially for the fair trade and eco-labeled coffee. The results are consistent with “desirability bias” (Bradburn, 1983), since those two labels “carry a strong identification with the working conditions and natural habitat of the country of origin of the African interviewer” (Loureiro and Lotade, 2005). However, since there are only two interviewers who certainly differ in dimensions other than race, it is difficult to draw any general conclusion regarding race-specific interviewer effects in

environmental valuation surveys. Two alternative approaches are to either conduct controlled experiments as in Bateman and Mawby (2004) or use the variation in an interviewer sample to statistically control for other interviewer characteristics besides the one of interest.

Our research applies the latter approach by analyzing data from 132 interviewers and 3,780 respondents to control for a number of respondent, interviewer and interview characteristics. In the process, we attempt to answer the following series of questions: Are there systematic gender and race-specific interviewer effects in environmental valuation surveys? Are there particular interviewer-respondent interactions that lead to significant biases? And finally, how can we control for the interviewer effects in environmental valuation surveys to improve the reliability and validity of the survey results?

The rest of the paper is structured as follows. Section 2 provides information on the survey method and data collection process. Section 3 reports the major survey results and tests both direct and interactive interviewer effects. Section 4 concludes with a summary and general discussion of the results.

## **2. Survey Method and Data Collection**

To answer the questions proposed at the end of the Introduction, we test for interviewer effects in a large-scale survey on recycling valuation and behavior. The survey topic, recycling, has attributes of a private good that directly benefit the household (e.g., cash for aluminum cans or lower fees for municipal waste service) and attributes of a public good that benefit society as a whole (e.g., reducing the flow of waste to landfills, extending the life of current landfills, reducing contamination of municipal water supplies, etc.). On the one hand, respondents may find it difficult to place a monetary value on the public, non-market benefits of recycling services and therefore may rely on the interviewer and the interview process to

aid them in converging on a value. On the other hand, because respondents are likely to have more experience with recycling than other environmental goods, any interviewer effects we find in this paper are expected to be a lower bound for surveys associated with less tangible environmental goods (e.g., clean air and water, mitigation of climate change, etc.).

We used a random-digit dialed (RDD) telephone survey. As noted by Arrow *et al.* (1993), “telephone interviews have some advantage in terms of cost and centralized supervision” as compared to face-to-face interviews. Those advantages are especially relevant in the current study because: 1) our budget was insufficient to conduct face-to-face interviews over the entire sampling region<sup>2</sup>; 2) centralized supervision was critical to minimize variance during interview process, which helped us control for factors that may interfere with the two interviewer characteristics of interest; and 3) the RDD mechanism allowed a large random sample of respondents with diversified characteristics, which is important for testing the interactions between interviewers and respondents.

In the winter of 2002, the Social and Economic Science Research Center (SESRC) at Washington State University (under the supervision of Drs. John Tarnai and Don Dillman) administered an RDD telephone survey regarding recycling behavior to nearly 4,000 households in 40 western U.S. cities with populations over 50,000.<sup>3</sup> The demographics of the 132 interviewers and details on the interview process are reported in Table 1. We will revisit those interviewer characteristic and interview process variables later in the paper.

The principal investigators and SESRC took a number of steps to ensure that responses from individuals were as accurate and reliable as possible. First, interviewers followed a

---

<sup>2</sup> The goal of the research project was to estimate the value of recycling services to the nation as a whole, but the funding agency recommended narrowing the region to the western United States. The western U.S. is likely to provide a lower bound (Arrow *et al.*, 1993) to the national benefits of recycling given fewer constraints on landfill space.

<sup>3</sup> The survey instrument, a list of the 40 cities in our sample, and information on the calculation of the response rate (27.5%) and the cooperation rate (49%) are available at [www.uwyo.edu/aadland/research/recycle/datareport.pdf](http://www.uwyo.edu/aadland/research/recycle/datareport.pdf). For further information on our survey design and econometric analysis, see Aadland and Caplan (2006).

script during the interview process. For example, at the beginning interviewers stated that “we would like to speak with an adult in your household who is responsible for managing your household’s garbage and/or recycling activities.” In this manner we were able to speak with the person in the household most knowledgeable regarding recycling. Second, interviewers were required to participate in a minimum of eight hours of basic interview training and an additional two hours of project-specific training (i.e., background information, purpose of the study, definitions, and a review of questions/content of the survey). Interviewers also practiced for a minimum of one hour on the CATI questionnaire before making actual calls. Third, interviewer performance was regularly monitored and measured to maintain data quality and continuity in the telephone data-collection process. SESRC states that “training and scoring are guided by two principles: 1) respondents should receive information that is delivered by the interviewer in an unbiased manner; and 2) every respondent should receive the same stimulus from each interviewer.” Training, monitoring and scoring are implemented to “minimize interviewer effects.” Fourth, respondents were informed at the beginning of the survey that “I am not selling anything and your answers are completely anonymous and voluntary. We also wish to assure you that this is strictly a scientific survey, which is not in any way associated with your state or local government.” Fifth, since respondents may have less time in a phone survey to understand the questions than in a face-to-face survey (Arrow *et al.*, 1993), the principal investigators monitored several of the 100 pretest interviews to ensure that respondents comprehended the questions and answered them in a reasonable fashion.

The survey is broken into three sections. The first section asks questions relating to the respondent’s knowledge and use of existing curbside and drop-off recycling services. The second section elicits the respondent’s maximum willingness to pay (WTP) for their curbside recycling program (CRP). Households located in a community without a CRP were instead



described a hypothetical program. The third and final section of the survey contains demographic questions.

As is common in valuation studies, rather than ask for an open-ended WTP value, the respondent was presented with a sequence of dichotomous-choice questions of whether to purchase recycling services at an offered price (often referred to as “bids” in the contingent-valuation literature). Previous research has found that respondents are generally more comfortable with this type of take-it-or-leave-it choice at a fixed price as it closely resembles the daily purchasing decisions they make for private goods (Mitchell and Carson, 1989). The first question was: “Would you be willing to pay \$ $b$  for the service?” The opening bid,  $b$ , was chosen randomly from a set of predetermined values.<sup>4</sup> Based on her response to the opening bid, the respondent was then asked a similar follow-up question, but with a larger bid value,  $2b$ , if she answered “yes” (i.e., she is willing to pay at least  $b$  for the service) or a smaller bid,  $0.5b$ , if she answered “no” (i.e., she is unwilling to pay  $b$  for the service). Based on the responses to the opening bid and follow-up questions, the respondent’s latent WTP may be placed in one of four regions:  $(0, 0.5b)$ ,  $(0.5b, b)$ ,  $(b, 2b)$  or  $(2b, \infty)$ .

### **3. Survey Results and Interviewer Effects**

#### **3.1 Interviewer Gender and Race Effects**

A critical question when testing for interviewer effects in telephone surveys is whether the respondent can perceive an interviewer’s characteristics. In a lab-conducted telephoned conversation (Lass *et al.*, 1979), respondents were able to identify the sexes of 99.2% of male speakers and 98.9% of female speakers. Data from the field are lower, but still indicate that most respondents can perceive the interviewer’s gender correctly. For example, Schejbal *et al.* (1993) report that 97% of respondents in a telephone survey correctly identified the

---

<sup>4</sup> The opening bids were chosen with equal probabilities from the set of integers two through 10. This set encompasses the range of household fees charged by the communities in our sample.

interviewer's gender. On race, Davis (1997) reports that the interviewee correctly identified the interviewers' race 79% of the time in a telephone survey in which there were both white and African American interviewers. Wilson and Olesen (2002) find that 74% of respondents from a national sample of U.S. adults were able to correctly identify the race of an African American interviewer over the phone, while white interviewers were correctly identified 83% of the time. Overall, previous research indicates that the majority of respondents can correctly identify the gender and race of the interviewer, although they are better at identifying gender than race.

For the reason mentioned above, we focus on the race and gender effect in our telephone survey of curbside recycling services. The *Social Attribution Model* predicts a direct effect of interviewer characteristics when respondents alter their answers to conform to the interviewers' norms and expectations. In the current survey, respondents may infer the interviewer's attitude on the environmental good (the recycling program) from the interviewer's race and gender and then alter their WTP accordingly.

Previous research has found robust gender and race differences in environmental issues. Females tend to be more environmental conscious and more likely to participate in green activities, such as recycling, than males (Diamantopoulos *et al.*, 2003). Several studies find that white respondents claim to have more concern for the environment than their African American counterparts (Baugh, 1991; Newell and Green, 2001). Given that interviewers' gender and race may imply different attitudes on environmental concerns, the *Social Attribution Model* predicts that respondents may sense a higher expected WTP when interviewed by a white or female interviewer and adjust their answers accordingly. Hence we propose two hypotheses regarding direct interviewer effects:

**H1** (Race Effect): White interviewers elicit higher WTP from respondents than non-white interviewers, all else equal.<sup>5</sup>

**H2** (Gender Effect): Female interviewers elicit higher WTP from respondents than male interviewers, all else equal.

To test these two hypotheses, we estimate a hierarchical regression model with clustering effects:

$$WTP_{ij} = X'_{ij}\gamma_X + Z'_{ij}\gamma_Z + \alpha_j + \varepsilon_{ij} \quad (1)$$

where  $WTP_{ij}$  is the midpoint of the WTP region<sup>6</sup> for household  $i$  and interviewer  $j$ ;  $X_{ij}$  is a vector of demographic variables (e.g., respondent's age, gender, employment status, education, years in community, race, and household income) and curbside and drop-off recycling attributes;  $Z_{ij}$  is a vector of interview variables (e.g., interviewer race, gender, and age; interview day-of-the-week and time-of-day),  $\gamma_X$  and  $\gamma_Z$  are constant parameter vectors;  $\alpha_j$  represents the interviewer-specific (clustering) random effect; and  $\varepsilon_{ij}$  is a random error term. The model is estimated using two-step feasible generalized least squares (Greene, 2001).

Table 2 reports the variable definitions and the estimation results. The econometric results imply that the average household is willing to pay \$3.97 per month for the recycling

---

<sup>5</sup> Unlike most previous research in which there were only white and African American interviewers, in the current study, there were a few Spanish and Asian interviewers, although the majority interviewers were either white or African American. Hence we use the white/non-white categorization instead of the usual white/African American one.

<sup>6</sup> Valuation surveys with a referendum WTP question (i.e., take-it-or-leave-it choice at a fixed price), typically estimate the model using discrete-choice econometric methods such as logit or probit. For transparency and because standard software packages do not offer clustering options in such models, we assign a single WTP estimate to each household. To be conservative, we choose the lower bound of the WTP interval as the point estimate for WTP. We experimented with other less conservative methods of assigning a WTP estimate. The results were qualitatively similar and are available upon request.

service. As shown in the table, the regression controls for a rich set of household and CRP attributes ( $X_{ij}$ ) that have been shown to be correlated with WTP (Blaine *et al.*, 2005; Hong *et al.*, 1993; Kinnaman, 2001; Lake *et al.*, 1996). In particular, female, white, and young respondents had a higher WTP for this environmental good than their male, non-white, and older counterparts.

Both **H1** and **H2** are supported by the data at a 5% significance level using a one-tailed  $t$  test. All else equal, respondents' monthly WTP was 7% (\$0.27) higher if interviewed by a white interviewer than a non-white interviewer [ $t(3,658)=1.85$ ,  $p<0.05$ ] and was 8% (\$0.32) higher if interviewed by a female interviewer than a male interviewer [ $t(3,658)=2.56$ ,  $p<0.01$ ].<sup>7</sup>

In addition to the race and gender effects, we also find that *when* the survey was completed had a significant influence on the stated WTP. Surveys completed on weekends had a lower WTP (\$0.56) than those completed on weekdays [ $t(3,658)=-3.00$ ,  $p<0.01$ ]. The time of day at which the survey was completed mattered too. All else equal, respondents who answered the survey during the daytime (9am-5pm) revealed lower monthly WTP than those who answered in the evenings (5pm-9pm). The WTP difference was \$0.50 [ $t(3,658)=-3.91$ ,  $p<0.01$ ]. For discussion purposes, we call those two differences "timing effects". To our knowledge, timing effects have not been reported in previous survey research. The timing effects stand up after controlling for a wide variety of respondent and interviewer characteristics, making it unlikely that the timing effects are related to individual differences. Instead, we believe that the observed timing effects are systematic and contribute to variation in WTP responses.

---

<sup>7</sup> The percentages are based on an average estimated household WTP of \$3.97 per month.

### 3.2 Interviewer-Respondent Interactions

Having a rich variety of interviewer and respondent characteristics, the data also allow us to test for the influence of interviewer-respondent interactions as proposed by the *Conditional Social Attribution Model*. Table 3 describes the survey numbers in each interactive category.

A common explanation for the interaction effects between interviewer and respondent is given by social distance theory (Williams 1964). Social distance theory implies that interviewer effects are smaller when there is a shorter social distance between the interviewer and the respondents, for example when both are of the same social status or demographic category. We measure social distance on two dimensions: whether the interviewer and the respondent were of the same gender or of the same race.

Following the race and gender effects reported in the last section, we use social distance theory to hypothesize the following conditional interviewer effects:

**H3a:** The Race Effect is smaller when both interviewer and respondent are of the same race.

**H3b:** The Gender Effect is smaller when both interviewer and respondent are of the same race.

**H4a:** The Race Effect is smaller when both interviewer and respondent are of the same gender.

**H4b:** The Gender Effect is smaller when both interviewer and respondent are of the same gender.

However, the *Conditional Social Attribution Model* is also consistent with an established propensity-to-conform differential across various types of respondents. These differences in the propensity to conform may cause some types of respondents to be more susceptible to interviewer effects. Since previous research finds that females are more likely

to conform than males when their actions are observable (Eagly, 1978), we hypothesize that female respondents are more susceptible to interviewer effects than male respondents. More specifically, we hypothesize that

**H5a:** The Race Effect is larger for female respondents than male respondents.

**H5b:** The Gender Effect is larger for female respondents than male respondents.

Note that **H4b** (from social distance theory) and **H5b** (from propensity-to-conform differentials) are conflicting hypotheses when considering female respondents. On the one hand, when a female respondent is interviewed by a female interviewer, she may feel less social pressure to conform than a male respondent because of the “shorter” social distance between the two. While on the other hand, a female respondent may be more susceptible to the interviewer gender effect than a male respondent because, all else equal, females may be more likely to conform to social norms than males.

To test the six hypotheses, we ran five separate regressions (**H4b** and **H5b** are tested in a single regression). Each regression is based on equation (1) with an additional interaction term. Table 4 reports the interaction coefficients and their *t* values. After using the Benjamain and Hochberg (1995) process to control for the false discovery rate<sup>8</sup>, two coefficients (**H5a** and **H5b**) are significant, both supporting the propensity-to-conform differential hypotheses that female respondents are generally more susceptible to the interviewer effects. The impact of interviewer race and gender were stronger for female respondents than for male respondents. For the race effect, the difference was \$0.95 per month [ $t(3,657)=3.20$ ,  $p<0.001$ ], and the difference was \$0.81 per month [ $t(3,657)=3.38$ ,  $p<0.001$ ] for the gender effect.

---

<sup>8</sup> We need to control for false discovery rate because we have multiple null hypotheses. Testing them independently may artificially inflate the rejection rate and lead to a false discovery.

Interestingly, none of the hypotheses predicted by social distance theory (*H3a* through *H4b*) were significant after controlling for false discovery. Note also that the data reject *H4b* (predicted by the social distance model) in favor of *H5b* (the propensity-to-conform differential hypothesis). Does this imply that the social distance theory is inappropriate for analyzing all interviewer-respondent interactions? Not necessarily. Instead, we argue that the interactions may depend on the topic. Take for example the interaction between interviewer gender and respondent gender. If the survey had instead been on a gender-sensitive topic, male respondents may have felt a greater social pressure (due to the social distance in gender) to give answers that would not offend a female interviewer. As a result, male respondents become more susceptible to the interviewer gender effect than female respondents do (consistent with *H4b*).

#### **4. General Discussion and Conclusions**

Interviewer effects have been found in a variety of topics and there is a general concern regarding the quality of survey data in the presence of such effects (Hox, 1994). In telephone surveys, race and gender of the interviewer are two primary sources of interviewer effects because respondents are often capable of identifying the interviewer's race and gender from voice cues and first names. The respondent may then infer the interviewer's attitude and norm on the survey topic and answer accordingly. It is common in the interview-effects literature to study the effects of gender and race separately (e.g. Cotter *et al.*, 1982; Groves and Fultz, 1985)<sup>9</sup>. Many times this is due to a lack of quality data.

In this paper, we test for both interviewer gender and race effects using a rich dataset produced from an environmental valuation telephone survey. The data are from a telephone survey in 40 cities across the western United States. The dataset has several advantages: (1)

---

<sup>9</sup> There are a few exceptions. For example, Callegaro *et al.* (2001) investigated both gender and race effects in an RDD telephone pre-election poll in Minneapolis. Their results show that the importance of including both variables in the analysis.

it is a large representative sample consisting of 3,780 respondents and 132 interviewers, (2) it includes a rich variety of interviewer and respondent characteristics (e.g., gender and race of both interviewers and respondents, and other demographic information as shown in Tables 1 and 2), and (3) it includes other relevant control variables (e.g., interview time and date) that are often excluded from other studies of interviewer effects. The richness of the dataset enables us to test for a host of possible interactions between interviewers and respondents.

To our knowledge, this is the first study testing for systematic gender and race effects in an environmental valuation survey. This is an important omission because the existence of between-interviewer variations inflates the total variance of respondents' answers (Tucker, 1983). Due to the non-market characteristics of many environmental goods, policy makers must rely on valuation surveys to gauge the society's value for various environmental policies. In addition, individuals often have less market experience with environmental goods and may be more prone to interviewer effects as they try to form and state a subjective value for the good.

After including a variety of control variables and accounting for interviewer clustering, we find that race and gender effects exist in the valuation of recycling programs. Respondents stated higher WTPs when interviewed by a white interviewer than by a non-white interviewer, and when interviewed by a female interviewer than by a male interviewer. These findings are consistent with both the *Social Attribute Model* and previous empirical results on gender and race differences for environmental concerns (Diamantopoulos *et al.*, 2003; Newell and Green, 2001).

In addition to the direct interviewer effects, we find significant interaction effects between interviewer and respondent characteristics, as predicted by the *Conditional Social Attribute Model*. For example, female respondents were more susceptible to gender and race effects than male respondents. These interactions tend to be consistent with gender



differences in the propensity to conform, as reported by social psychologists (Eagly, 1978; Baron *et al.*, 1992), but are not consistent with social distance theory which is commonly applied when studying interviewer-respondent interactions.

Finally, we also find “timing” has a significant effect on the valuation estimates. All else equal, surveys completed on weekends are associated with lower WTPs than those that were completed on weekdays; while those completed during the daytime (9am-5pm) revealed lower WTPs than those in the evenings (5pm-9pm). To our knowledge, these survey “timing effects” are new to the interview-effects literature.

There are two primary contributions of our research. First, we provide evidence on the existence of systematic gender and race effect and timing effect in an environmental valuation survey, even after controlling for a variety of respondent characteristics and the clustering of interviewers. If the interview design and implementation is skewed toward a particular type of interviewer or at a particular time of day, the valuation estimates resulting may be significantly biased and lead to an inefficient allocation of public resources. Second, our results provide evidence in favor of the propensity-to-conform differential hypothesis in explaining the presence of interview interaction effects between respondents and interviewers. The significance of this finding may well go beyond interviewer effects in valuation surveys and make a contribution to the more general interviewer-effect literature in the social sciences.

There are a couple of notable limitations of the current study. First, the survey was conducted through the telephone. Interviewer effects are usually smaller in telephone surveys than in face-to-face surveys (Groves and Kahn, 1979). Hence, the interviewer effects in this telephone survey only provide a lower bound for those in face-to-face surveys. Given face-to-face interviews are the preferred approach in valuation surveys, it would be interesting to conduct a comparable interviewer-effect analysis using the face-to-face method.

Another possibility would be a carefully designed laboratory experiment to test for interviewer effects. For example, one could choose two interviewers who are similar in all respects other than gender (e.g., they could be of the same race, age, education, experience and be trained to use the same transcription in similar accent and intonation). A good example of using experimental manipulation to study interviewer effects in a face-to-face valuation survey is Bateman and Mawby (2004), although they did not investigate gender or race.

Another limitation of the current study is that it does not examine interviewer age effects, which is one of the more common effects in the literature (e.g., Norris and Hatcher, 1994). Although we had information on interviewers' age, we suspected that respondents had difficulty in identifying the interviewer's age over the phone. Instead we treated interviewer age as one of many control variables.

We conclude by discussing the practical implications of our paper. In an ideal world, interviewer effects could be reduced by designing and implementing surveys with an even distribution of characteristics for the respondent, interviewer, and interview process. For example, interviews could be uniformly distributed across various times-of-the-day, day-of-the-week; as well as interviewer race and gender. This type of interview variation would tend to mitigate any systematic biases associated with relying on a single type of interview (e.g., only weekdays or only female interviewers). In practice, however, logistical consideration often trump concerns for interview bias. Interviewer hiring and survey administration practice are often under a wide range of constraints imposed by labor market and organizational efficiencies. Therefore, common practice is to only be concerned about interviewer effects for opinion surveys on socially sensitive topics (e.g., domestic violence, sexual attitudes, race relations, etc.). Thus, although we suspect that race, gender and timing interviewer effects are likely to inject systematic biases in environmental valuation surveys;

we hesitate to suggest any radical changes in survey design based on the results from one study. Further research is needed to test the generality and nature of interviewer effects in environmental valuation surveys.

That said, the evidence of interviewer effects in the current study should not be ignored, especially given the large sample size (3,780 surveys) and the magnitude of the effects (e.g. the 14% WTP weekend interviewer effect and the 8% WTP gender effect). Instead, researchers and policy makers who rely on environmental valuation surveys should be made aware of these possible systematic interviewer biases. They can then include the characteristics of the interviewers and interview process as control variables and interpret the resulting valuation estimates accordingly.

We thank John Tarnei, Kent Miller, and the Washington State University Survey Research Laboratory for conducting the survey for this study. We also thank Jonathan Baron, Arthur Caplan, Burke Grandjean, Howard Kunreuther, and participants at the Psychology beer seminar, University of Pennsylvania for providing constructive criticism on earlier versions of the paper.

## References

- Aadland, D. and A. Caplan. 2006. "Curbside Recycling: Waste Resource or Waste of Resources?" *Journal of Policy Analysis and Management*, 25(4), 855-874.
- Aadland, D., A. Caplan and O. Phillips. 2007. "A Bayesian Examination of Information and Uncertainty in Contingent Valuation." *Journal of Risk and Uncertainty*, 35(2), 149-178.
- Arrow, K., R. Solow, P. Portney, E. Leamer, R. Radner and H. Shuman. 1993. *Report of the NOAA Panel on Contingent Valuation*. Available at: [www.darp.noaa.gov/library/pdf/cvblue.pdf](http://www.darp.noaa.gov/library/pdf/cvblue.pdf).
- Barath, A. and C. Cannell. 1976. "Effect of Interviewer's Voice Intonation." *Public Opinion Quarterly*, 40, 370-373.
- Baron, R., N. Kerr and N. Miller. 1992. *Group Process, Group Decision, Group Action*. Pacific Grove, CA: Brooks Cole.
- Bateman, I. and J. Mawby. 2004. "First Impressions Count: A Study of the Interaction of Interviewer Appearance and Information Effects in Contingent Valuation Studies." *Ecological Economics*, 49(1), 47-55.
- Baugh, J. 1991. "African-Americans and the Environment: A Review Essay." *Policy Studies Journal*, 19(2), 182-191.
- Benjamini, Y. and Y. Hochberg. 1995. "Controlling the False Discovery Rate—A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289-300.
- Blaine, T., F. Lichtkoppler, K. Jones, and R. Zondag. 2005. "An Assessment of Household Willingness to Pay for Curbside Recycling: A Comparison of Payment Card and Referendum Approaches." *Journal of Environmental Management*, 76, 15-22.
- Boyle, K. and R. Bishop. 1988. "Welfare Measurement Using Contingent Valuation: A Comparison of Techniques." *American Journal of Agricultural Economics*, 70, 20-28.
- Bradburn, N. 1983. "Response effects", in Rossi, P.H., Wright, J.D., Anderson, A.B. (Eds.), *Handbook of Survey Research*, Academic Press, New York, NY.
- Brookshire, D., M. Thayer, W. Schulze, and R. d'Arge. 1982. "Valuing Public Goods: A Comparison of Survey and Hedonic Approaches," *American Economic Review*, 72, 165-177.
- Callegaro, M., F. De Keulenaer, J. Krosnick, and R. Daves. 2006. "Interviewer Effect in an RDD Telephone Pre-election Poll in Minneapolis 2001. An Analysis of the Race and Gender Interviewer Effect." In the 2005 Proceedings of the American Statistical Association, 60th Annual Conference of the American Association for Public Opinion Research [CD-ROM] (pp. 3815-3821). Alexandria, VA: American Statistical Association.

- Carson, R. and R. Mitchell. 1993. "The Value of Clean Water: The Public's Willingness to Pay for Boatable, Fishable and Swimmable Quality Water." *Water Resources Research*, 29, 2445-2454.
- Carson, R., R. Mitchell, W. Hanemann, R. Kopp, S. Presser, and P. Ruud. 1992. *A Contingent Valuation Study of Lost Passive Use Values Resulting from the Exxon Valdez Oil Spill*, report to the Attorney General of Alaska, Natural Resource Damage Assessment, Inc., La Jolla, CA.
- Clarke, P., K. Sproston, and R. Thomas. 2003. "An Investigation into Expectation-Led Interviewer Effects in Health Surveys." *Social Science & Medicine*, 56, 2221-2228.
- Cotter, P., J. Cohen, and P. Coulter. 1982. "Race-of-Interviewer Effects in Telephone Interviews." *Public Opinion Quarterly*, 46, 278-84.
- Davis, D. 1997. "Nonrandom Measurement Error and Race of Interviewer Effects among African Americans." *Public Opinion Quarterly*, 61(1), 183-207.
- Desvousges, W., V. Smith, and A. Fisher. 1987. "Option Price Estimates for Water Quality Improvements: A Contingent Valuation Study for the Monongahela River." *Journal of Environmental Economics and Management*, 14, 248-267.
- Diamantopoulos, A., B. Schlegelmilch, R. Sinkovics, G. Bohlen. 2003. "Can Socio-demographics Still Play a Role in Profiling Green Consumers? A Review of the Evidence and an Empirical Investigation." *Journal of Business Research*, 56, 465-480.
- Dijkstra, W. 1987. "Interviewing Style and Respondent Behavior: An Experimental Study of the Survey-Interview." *Sociological Methods & Research*, 16, 309-34.
- Eagly, A. 1978. "Sex Differences in Influenceability." *Psychological Bulletin*, 85, 86-116.
- Fendrich, M., T. Johnson, C. Shalligram and J. Wislar. 1999. "The Impact of Interviewer Characteristics on Drug Use Reporting by Male Juvenile Arrestees." *Journal of Drug Issues*, 29(1), 37-58.
- Finkel, S., T. Guterbock and M. Borg. 1991. "Race-of-Interviewer Effects in Pre-election Poll: Virginia 1989." *Public Opinion Quarterly*, 55(3), 313-330.
- Greene, W. 2001. *Econometric Analysis*, 5th ed. Upper Saddle River, NJ: Prentice Hall.
- Groves, R., and R. Kahn. 1979. *Surveys by Telephone: A National Comparison with Personal Interviews*. New York: Academic Press.
- Groves, R. and N. Fultz. 1985. "Gender Effects among Telephone Interviewers in a Survey of Economic Attitudes." *Sociological Methods & Research*, 14, 31-52.
- Hersch, J. and W. Viscusi. 2006. "The Generational Divide in Support for Environmental Policies: European Evidence." *Climatic Change*, 77, 121-136.

- Hong, S., R. Adams, and A. Love. 1993. "An Economic Analysis of Household Recycling of Solid Wastes: The Case of Portland, Oregon." *Journal of Environmental Economics and Management*, 25, 136-146.
- Hox, J. 1994. "Hierarchical Regression Models for Interviewer and Respondent Effects." *Sociological Methods Research*, 22(3), 300-318.
- Johnson, T., M. Fendrich and C. Shaligram. 2000. "An Evaluation of the Effects of Interviewer Characteristics in an RDD Telephone Survey of Drug Use." *Journal of Drug Issues*, 30(1), 77-101.
- Kinnaman, T. 2001. "Explaining Household Demand for the Collection of Solid Waste and Recycling" in D. Fullerton and T. Kinnaman (Eds). *The Economics of Household Garbage and Recycling Behavior*, Cheltenham, UK: Edward Elgar.
- Lake, I., I. Bateman and J. Parfitt. 1996. "Assessing a Kerbside Recycling Scheme: A Quantitative and Willingness to Pay Case Study." *Journal of Environmental Economics and Management*, 46, 239-54.
- Landis, J., D. Sullivan and J. Sheley. 1973. "Feminist Attitudes as Related to Sex of the Interviewer." *Pacific Sociological Review*, 16, 305-314.
- Lass, N., J. Tecca, R. Mancuso and W. Black. 1979. "The Effect of Phonetic Complexity on Speaker Race and Sex Identification." *Journal of Phonetics*, 5, 105-118.
- Lee, T. 2001. "Language-of-Interviewer Effects and Latino Mass Opinion." Working paper, John F. Kennedy School of Government, Harvard University.
- Loomis, J., L. Ellingson, A. González-Cabán, and A. Seidl. 2006. "The Role of Ethnicity and Language in Contingent Valuation Analysis: A Fire Prevention Policy Application." *American Journal of Economics and Sociology*, 65(3), 559-586.
- Loureiro M., and J. Lotade. 2005. "Interviewer Effects on the Valuation of Goods with Ethical and Environmental Attributes." *Journal of Environmental and Resource Economics*, 30(1), 49-72.
- McFadden, D. "Economic Choices." *American Economic Review*, 91(3), 351-378.
- McKenzie, J. 1977. "An Investigation into Interviewer Effects in Market Research." *Journal of Marketing Research*, 14, 330-336.
- Mitchell, R. and R. Carson. 1989. *Using Surveys to Value Public Goods: The Contingent Valuation Method*, Resources for the Future, Washington, D.C.
- Newell, S. and C. Green. 1997. "Racial Differences in Consumer Environmental Concern." *Journal of Consumer Affairs*, 31(1), 53-69.
- Norris, D. and J. Hatcher. 1994. "The Impact of Interviewer Characteristics on Response in a National Survey of Violence Against Women." *Proceedings of the Survey Research Methods Section*. American Statistics Association ([www.amstat.org/sections/srms/Proceedings/](http://www.amstat.org/sections/srms/Proceedings/)).

- Schejbal, J., H. Sachs and P. Lavrakas. 1993. "Hello Do You Remember Us? What Respondents Remember about the Interview." Paper presented at the Annual Conference of the Midwest Association for Public Opinion Research (MAPOR), Chicago, IL.
- Singer, E., M. Frankel, and M. Glassman. 1983. "The Effect of Interviewer Characteristics and Expectations on Response." *Public Opinion Quarterly*, 47, 68-83.
- Tucker, C. 1983. "Interviewer Effects in Telephone Surveys." *Public Opinion Quarterly*, 47, 84-95.
- Udow, A. 1942. "The 'Interviewer-Effect' in Public Opinion and Market Research Surveys." *Archives of Psychology* (Columbia University) 277.
- Walsh, R., R. Bjonback, R. Aiken, and D. Rosenthal. 1990. "Estimating the Public Benefits of Protecting Forest Quality." *Journal of Environmental Management*, 30, 175-189.
- Williams Jr., J. 1964. "Interviewer-Respondent Interaction: A Study of Bias in the Information Interview." *Sociometry*, 27, 338-352.
- Williams Jr., J. 1968. "Interviewer Role Performance: A Future Note on Bias in the Information Interview." *Public Opinion Quarterly*, 32, 287-294.
- Wilson, D. and E. Olesen. 2002. "Perceived Race of Interviewer Effects in Telephone Interviews." Paper presented at the 57th AAPOR/WAPOR conference, St. Pete Beach, FL.

Table 1: Descriptive Statistics of the Survey

Total number of surveys	3780								
Total number of interviewers	132								
Surveys completed per interviewer	Mean	Median			Min		Max		
	29	25			1		97		
Interviewer characteristics	Race (number of interviewers)		Gender (number of interviewers)			Age (years)			
	White	Non-white	Male	Female	Mean	Median	Min	Max	
	101	31	36	96	21	21	16	51	
Interview process	Date (number of surveys)		Time (number of surveys)			Length (minutes)			
	Weekday	Weekend	Early Morning	Day Time	Evening	Mean	Median	Min	Max
	3121	659	41	1423	2316	9	8	4	35



Table 2. Control and Interview Variables in the General Model (N = 3780)

Variables	Mean	Coef.	Description
Ethical Duty	0.87	2.31***	Do you feel an ethical duty to recycle to help the environment? 1=yes, 0=no.
Monetary	0.48	-0.02	Are you motivated to recycle in order to save money? 1=yes, 0=no.
Primarily Ethics	0.24	1.55***	Which most encourages you to recycle? 1=ethical duty, 0=save money.
Dropoff Distance	2.80	0.04*	Distance in miles to the nearest dropoff site.
Dropoff User	0.29	0.21	In the past 12 months has your household used dropoff recycling? 1=yes, 0=no.
Young	0.33	1.65***	1 if 18<Age<35, 0 otherwise.
Old	0.12	0.01	1 if 65<Age, 0 otherwise.
Male	0.41	-0.53***	1=male, 0=female.
High School	0.13	1.04**	Highest level of education in household? 1=high school graduate, 0=otherwise.
College Degree	0.58	1.51***	1=Associates or Bachelors degree, 0=otherwise.
Advanced Degree	0.25	2.40***	1=Masters, Ph.D. or equivalent professional degree, 0=otherwise.
Household Size	1.09	0.45***	Number of adults in household, other than the respondent.
Environment. Org.	0.09	0.64***	Anyone in your household belong to an environmental org.? 1=yes, 0=no.
Med Income	0.35	0.04	1 if \$35K/yr<Household Income<\$75K/yr, 0 otherwise.
High Income	0.32	0.27	1 if \$75K/yr<Household Income, 0 otherwise.
Employed	0.81	1.31	Adult with the highest income currently employed? 1=yes, 0=no.
Retired	0.12	-0.11	Adult with the highest income currently retired? 1=yes, 0=no.
Sorting Required	0.42	-0.001	1=CRP requires some sorting of recyclable materials, 0 otherwise.
Landfill Visit	0.52	-0.69***	Has anyone in your household visited your community's landfill? 1=yes, 0=no.
Landfill Distance	7.59	0.24	Distance to nearest landfill in miles.
Landfill Dist.>2mi.	6.91	-0.25	Distance above and beyond 2 miles to nearest landfill, 0 otherwise.
Hypothetical	0.47	1.14***	1=respondent valued a hypothetical CRP, 0=otherwise.
Precision	84.8	-0.02***	On a scale of 0-100, how certain are you of the answers to your WTP questions?
English	0.98	0.29	Is English your first language? 1=yes, 0=no.
Employer Recycle	0.50	-0.18	Do you recycle at work? 1=yes, 0=no.
Caucasian	0.78	0.44**	What racial group best describes you? 1=Caucasian, 0 otherwise.
Hispanic	0.08	0.68*	What racial group best describes you? 1=Hispanic, 0 otherwise.
African American	0.03	-0.002	What racial group best describes you? 1=African American, 0 otherwise.
Generation Link	0.38	-0.14	Were you (or others in your house) raised in recycling households? 1=yes, 0=no.
Neighbor Recycle	0.39	-0.86***	Do most of your neighbors currently recycle? 1=yes, 0=no.
Years in Comm.	15.67	-0.02***	How many years have you lived in your community?
Number of children	0.86	0.07	How many children under the age of 18 currently live in your home?
Fee Known	0.26	-0.37**	Respondent knows current CRP fee? 1=yes, 0=no.
Fee Difference	1.19	0.16***	Stated CRP fee minus actual CRP fee.
CRP Performance	0.29	0.56*	Job performance of your current CRP? 1=excellent or good, 0=fair or poor.
Short Cheap Talk	0.34	-0.09	1=received short cheap-talk statement, 0 otherwise.
Long Cheap Talk	0.34	0.02	1=received longer cheap-talk statement, 0 otherwise.
Bid	5.51	0.20***	Opening Bid ( <i>b</i> )
Interviewer age	0.57	0.23	1=interviewer age >=21 (average age), 0 otherwise.
Few Attempts	0.69	0.48***	Respondent available for survey after first or second dialing attempt.
Refusal	0.11	-0.31	1=if refusal for first call attempt, 0 otherwise.
Early Morning	0.01	0.80	1=if the interview occurred between 8 am and 9 am, 0 otherwise.
Evening	0.61	0.50***	1=if the interview occurred after 5 pm, 0 otherwise.
Weekend	0.12	-0.56***	1=if the interview occurred on the weekend, 0 otherwise.
White Interviewer	0.75	0.27**	1=Caucasian interviewer, 0 otherwise.
Female Interviewer	0.63	0.32***	1=female interviewer, 0 otherwise.

Notes. The description does not always exactly match the wording in the survey instrument. To see the exact wording and complete descriptive statistics, please refer to [www.uwyo.edu/aadland/research/recycle/datareport.pdf](http://www.uwyo.edu/aadland/research/recycle/datareport.pdf). In calculating the means, the relevant sample size is N = 3780. However, due to the nature of some variables (e.g., Dropoff Distance and Primarily Ethics) the mean is calculated using only the relevant subsample of respondents. City and missing-value dummy variables were included in the regression but not listed in the Table. We did not find any significant pair-wise correlation (i.e., coefficient > 0.80) between explanatory variables. (\*\*\*), (\*\*), and (\*) refer to statistical significance and 1, 5 and 10 percent levels respectively.

Table 3: Number of Surveys in Each Category

		<b>Interviewer</b>			
		<b>Race</b>		<b>Gender</b>	
<b>Respondent</b>		White	Non-white	Male	Female
<b>Race</b>	White	2266	692	1073	1885
	Non-white	580	242	307	515
<b>Gender</b>	Male	1166	374	549	991
	Female	1680	560	831	1409

Table 4: Interviewer-Respondent Interactions

Hypothesis	Coefficient	<i>t</i> value (df=3657 <sup>††</sup> )	p value (one-tail <i>t</i> test)	5% significant after controlling for FDR <sup>†</sup>
<b>H3a</b>	-0.29	-1.77	0.04	No
<b>H3b</b>	-0.10	-0.52	0.30	No
<b>H4a</b>	0.25	1.79	0.04	No
<b>H4b &amp; H5b</b>	0.81	3.20	0.001 <sup>†††</sup>	Yes (H5b)
<b>H5a</b>	0.95	3.38	0.000	Yes

Notes. <sup>†</sup>We use the Benjamain and Hochberg (1995) process to control for the false discovery rate with a 5% cutoff *q* value equal to 0.001. <sup>††</sup> In each regression, we include a set of control variables most of which are reported in Table 2. <sup>†††</sup>Two-tailed *t* test as **H4b** and **H5b** make opposite predictions.