Do People Mean What They Say? Implications for Subjective Survey Data

By MARIANNE BERTRAND AND SENDHIL MULLAINATHAN*

Many surveys contain a wealth of subjective questions that are at first glance rather exciting. Examples include: "How important is leisure time to you?"; "How satisfied are you with yourself?"; or "How satisfied are you with your work?" Yet despite easy availability, this is one data source that economists rarely use. In fact, the unwillingness to rely on such questions marks an important divide between economists and other social scientists.

This neglect does not come from disinterest. Most economists would probably agree that the variables these questions attempt to uncover are interesting and important; but they doubt whether these questions elicit meaningful answers. These doubts are, however, based on a priori skepticism rather than on evidence. This ignores a large body of experimental and empirical work that has investigated the meaningfulness of answers to these questions. Our primary objective in this paper is to summarize this literature for an audience of economists, thereby turning a vague implicit distrust into an explicit position grounded in facts. Having summarized the findings, we integrate them into a measurement-error framework so as to understand what they imply for empirical research relying on subjective data. Finally, in order to calibrate the extent of the measurement-error problem, we perform some simple empirical work using specific subjective questions.

I. Evidence on Subjective Questions

A. *Cognitive Problems*

We begin by summarizing the experimental evidence on how cognitive factors affect the way people answer survey questions.[1] A set of experiments has shown that simple manipulations can affect how people process and interpret questions. One first interesting manipulation comes from the ordering of questions: whether question X is preceded by question Y or vice versa can substantially affect answers. One reason for this ordering effect is that people attempt to provide answers consistent with the ones they have already given in the survey. A second issue is that prior questions may elicit certain memories or attitudes, which then influence later answers. In a striking study, respondents were asked two happiness questions: "How happy are you with life in general?" and "How often do you normally go out on a date?" When the dating question came first, the answers to both were highly correlated, but when it came second, they were basically uncorrelated. Apparently, the dating question induced people to focus on one aspect of their life, an aspect that had undue effects on their subsequent answer.

Another cognitive effect is the importance of question wording. In one classic example, researchers compared responses to two questions: "Do you think the United States should forbid public speeches against democracy?" and "Do

* Bertrand: Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, IL 60637, NBER, and CEPR; Mullainathan: Department of Economics, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02173, and NBER.

[1] Due to space constraints, we will just mention two books that are good sources for reviews of the experimental evidence: Judith M. Tanur (1992) and Seymour Sudman et al. (1996). A fuller list of references is available in the full version of this paper (Bertrand and Mullainathan, 2000).

you think that the United States should allow public speeches against democracy?" While more than half of the respondents stated that yes, public speeches should be "forbidden," three-quarters answered that no, public speeches should not be "allowed." Evidence of such wording effects are extremely common.

Cognitive problems also arise due to the scales presented to people. In an experiment, German respondents were asked how many hours of television they were watching per day. Half of the respondents were given a scale that began with ≤30 minutes and then proceeded in half-hour increments ending with 4.5+ hours. The other respondents were given the same scale except the first five answers were compressed so that it began with ≤2.5 hours. Only 16 percent of the respondents given the first set of response alternatives reported watching more than 2.5 hours of TV per day, but 32 percent of the respondents given the second set of response alternatives reported watching more than 2.5 hours of TV per day. Respondents thus appear to be inferring "normal" TV viewing from the scale. The first scale, with a finer partition in the 0–2 hours range, suggests to subjects that this amount of TV-viewing is common. In fact, stating that the survey's purpose is to estimate the amount of TV-viewing greatly diminishes the scale effect.

An even more fundamental problem is that respondents may make little mental effort in answering the question, such as by not attempting to recall all the relevant information or by not reading through the whole list of alternative responses. As a consequence, the ordering of response alternatives provided matters, since subjects may simply pick the first or last available alternatives in a list. In the General Social Survey, for example, respondents are asked to list the most and least desirable qualities that a child may have out of a list of 13 qualities. Researchers surveyed people and gave them this list in either the GSS order or in reverse order. They found that subjects would rate the first or last listed qualities, whatever they were, as most important.

B. *Social Desirability*

Beyond purely cognitive issues, the social nature of the survey procedure also appears to play a large role in shaping answers to subjective questioning. Respondents want to avoid looking bad in front of the interviewer. A famous example is that roughly 25 percent of nonvoters report having voted immediately after an election. This overreporting is strongest among those who value norms of political participation the most and those who originally intended on voting. Other studies have noted that if one adds to a voting question a qualifier that "Many people do not vote because something unexpectedly arose ... ," the discrepancy rate between self-reported voting and actual voting drops.

Another example can be found in the self-reporting of racial attitude. Much evidence suggests that people are unwilling to report prejudice. For example, reported prejudice increases when respondents believe they are being psychologically monitored for truth-telling and decreases when the survey is administered by a black person.

C. *Non-Attitudes, Wrong Attitudes, and Soft Attitudes*

Perhaps the most devastating problem with subjective questions, however, is the possibility that attitudes may not "exist" in a coherent form. A first indication of such problems is that measured attitudes are quite unstable over time. For example, in two surveys spaced a few months apart, the same subjects were asked about their views on government spending. Amazingly, 55 percent of the subjects reported different answers. Such low correlations at high frequencies are quite representative.

Part of the problem comes from respondents' reluctance to admit lack of an attitude. Simply because the surveyor is asking the question, respondents believe that they should have an opinion about it. For example, researchers have shown that large minorities would respond to questions about obscure or even fictitious issues, such as providing opinions on countries that do not exist.

A second, more profound, problem is that people may often be wrong about their "attitudes." People may not really be good at forecasting their behavior or understanding why they did what they did. In a well-known experiment, subjects are placed in a room where two

ropes are hanging from the ceiling and are asked to tie the two ropes together. The two ropes are sufficiently far apart than one cannot merely grab one by the hand and then grab the other one. With no other information, few of the subjects are able to solve the problem. In a treatment group, the experimenter accidentally bumps into one of the ropes, setting it swinging. Many more people solve the problem in this case: subjects now see that they can set the ropes swinging and grab on an upward arc. Yet when they are debriefed and asked how they solved the problem, few of the subjects recognize that it was the jostling by the experimenter that led them to the solution.

A final and related problem is cognitive dissonance. Subjects may report (and even feel) attitudes that are consistent with their behavior and past attitudes. In one experiment, individuals are asked to perform a tedious task and then paid either very little or a lot for it. When asked afterwards how they liked the task, those who are paid very little report *greater* enjoyment. They likely reason to themselves, "If I did not enjoy the task, why would I have done it for nothing?" Rather than admit that they should just have told the experimenter that they were leaving, they prefer to think that the task was actually interesting. In this case, behavior shapes attitudes, and not the other way around.

II. A Measurement-Error Perspective

What do these findings imply for statistical work using subjective data? We will adopt a measurement-error perspective and assume that reported attitudes equal true attitudes plus some error term, $A = A^* + \varepsilon$. Statistically, we readily understand the case where ε is white noise. The above evidence, however, suggests two important ways in which the measurement error in attitude questions will be more than white noise. First, the mean of the error term will not necessarily be zero within a survey. For example, the fact that a survey uses "forbid" rather than "allow" in a question will affect answers. Second, many of the findings in the literature suggest that the error term will be correlated with observable and unobservable characteristics of the individual. For example, the misreporting of voting is higher in certain demographic groups (e.g., those who place more social value on voting).

There are two types of analysis that can be performed with subjective variables: using attitudes to explain behavior or explaining attitudes themselves. We will examine how mismeasurement affects both types of analyses. First, suppose that we are interested in using self-reported attitudes to explain behavior. Specifically, suppose that we estimate $Y_{it} = a + bX_{it} + cA_{it}$, while the true model is $Y_{it} = \alpha + \beta X_{it} + \gamma A^*_{it} + \delta Z_{it}$, where i represents individuals, t represents time, Y represents an outcome of interest, X represents observable characteristics, Z represents unobservable characteristics, and we assume for simplicity that Z is orthogonal to X. How will the estimated coefficient $\hat{c}$ compare to γ, given what we have learned about measurement error in attitude questions?

White noise in the measurement of A will produce an attenuation bias (i.e., a bias toward zero). The first measurement problem listed above, a survey fixed effect, will produce no bias as long as the appropriate controls (such as year- or survey-specific dummies) are included. The second problem, correlation with individual characteristics X and Z, will create a bias: $\hat{c}$ will now include both the true effect of attitude and the fact that the measurement error in A is correlated with unobservables. Hence, assuming that measurement-error problems are not dominant, subjective variables can be useful as control variables, but care must be taken in interpreting them. The estimated coefficient does not only capture the effect of attitude, but also the effect of other variables that influence how the attitude is self-reported. This is closely related to the causality problem that we often encounter, even with perfectly measured variables.[2]

We now turn to the second type of analysis, in which we are attempting to explain attitudes themselves. For example, we might ask whether high work hours increase loneliness. Specifically,

[2] An extreme example of this occurs when the measurement error is correlated with the variable of interest itself, as is suggested by cognitive dissonance. For example, people may report a lower preference for money if they are making less money. This is a case of pure reverse causation.

suppose that we estimate $A_{it} = a + bX_{it} + \varepsilon$, while the true model is $A^*_{it} = \alpha + \beta X_{it} + \gamma Z_{it}$.

In this setup, the white noise in the measurement of attitudes no longer causes bias. But the other biases now play a much more important role. Specifically, the fact that measurement error is correlated with individual characteristics will now severely bias X. For example, suppose that we see that those from rich backgrounds have a greater preference for money. As noted earlier, this might simply reflect the fact that a rich background affects the *reporting* of the preference for money. Such a correlation could thus be purely spurious. Notice that this problem is far more severe than in the previous analysis. First, the fact that an X helps predict "attitude" means very little if it is only predicting the measurement error in attitude. Therefore, one cannot argue as before that simply helping to predict is a good thing, irrespective of causality. Second, this is a problem that is much harder to solve than an omitted-variable bias problem. For example, it is hard to see how an instrumental variable could resolve this issue. One would need an instrument that affects X but not the measurement of attitude. But the above evidence tells us that X will likely affect measurement in a *causal* sense. This makes it very unlikely that such an instrument could be found in most contexts.

To summarize, interpreting the experimental evidence in a measurement-error framework provides two important insights. First, if the measurement error is small enough, subjective measures may be helpful as independent variables in predicting outcomes, with the caveat that the coefficients must be interpreted with care. Second, subjective variables cannot reasonably be used as dependent variables, given that the measurement error likely correlates in a very causal way with the explanatory variables.

III. How Much Noise Is There?

This leaves the important quantitative question: How much white noise error is there in the subjective questions we might be interested in? Can we in fact gain anything by adding responses to subjective questions to our econometric models?

TABLE 1—EFFECT OF ATTITUDE QUESTIONS ON FUTURE OUTCOMES

	Additional controls			
Question	(i)	(ii)	(iii)	(iv)
A. *Dependent Variable = Log Wage:*				
Value work?	0.08	0.07	0.07	0.06
	(0.03)	(0.03)	(0.02)	(0.02)
Value money?	0.08	0.08	0.05	−0.015
	(0.02)	(0.02)	(0.02)	(0.02)
Value steady job?	0.13	0.12	0.12	0.03
	(0.02)	(0.02)	(0.02)	(0.02)
Value family?	0.07	0.06	0.03	0.003
	(0.02)	(0.02)	(0.02)	(0.02)
Value friends?	−0.01	−0.01	0.02	0.00
	(0.02)	(0.02)	(0.02)	(0.02)
Value leisure?	0.09	0.08	0.06	0.02
	(0.02)	(0.02)	(0.02)	(0.02)
Value social causes?	−0.08	−0.08	−0.06	−0.01
	(0.02)	(0.02)	(0.02)	(0.02)
Positive toward self?	0.07	0.06	0.05	−0.00
	(0.02)	(0.02)	(0.02)	(0.02)
Reservation wage?	0.17	0.17	0.14	−0.00
	(0.03)	(0.03)	(0.03)	(0.02)
B. *Dependent Variable = Stayer:*				
Satisfied with Job?	0.08	0.08	0.08	—
	(0.01)	(0.01)	(0.01)	

Notes: Additional controls: (i) demographics; (ii) demographics plus family background; (iii) demographics plus family background plus log wage in 1983; (iv) person fixed effects. Demographic characteristics include education, sex, and race. Family background characteristics include father's education, mother's education, and family income in senior year (seven categories). "Stayer" is a dummy variable which equals 1 if there is no job change between the second and third follow-up. Each cell corresponds to a separate regression. Standard errors are in parentheses. Except in column (iv), outcomes are from the third follow-up survey, and attitudes are from the second follow-up. Column (iv) report panel regressions on all available survey periods. The regressions in column (iv) also include survey fixed effects.

To assess this, we turn to the "High School and Beyond" Senior Sample, which surveyed seniors in school in 1980 and then followed them every two years until 1986. This sample provides us with a set of subjective and objective variables in each of these waves.

In the first eight rows of Table 1, we correlate answers to a set of attitude variables with future income (thereby removing mechanical correlations with current income). Each cell in the

table corresponds to a separate regression. The dependent variable is log(salary) in 1985. In column (i), we add as controls the sex, race, and educational attainment of the respondent. Answers to the subjective questions clearly help predict individual income. The set of correlations is very intuitive. People who value money or a steady job more earn more. People who value social goals such as correcting inequalities around them earn less. People who have a positive attitude toward themselves earn more. Maybe somewhat intriguing, we find that people who care about their family earn substantially more. Even more intriguing, people who value leisure time also earn more. The second row shows that respondents' attitudes do not simply proxy for objective family background characteristics. Controlling for parents' education and family income in the senior year does not weaken the predictive power of the attitude variables. In column (iii), we show that attitude questions stay predictive of future income even after one controls for current individual income.

As a whole, these results suggest that noise does not dominate the measurement of these subjective questions. Attitudes actually predict income even beyond past income and background characteristics. Of course, we are not arguing for causality, merely that attitude variables add explanatory power.

Finally, one might wonder to what extent these variables are conveying any information beyond fixed individual characteristics. In column (iv), we exploit the panel nature of the High School and Beyond survey. We rerun the standard regressions with lagged attitude measures but also add person fixed effects. Most of the effects previously discussed disappear, except for the importance of work and the importance of having a steady job (which are now marginally significant). It therefore does not appear that changes in attitudes have as much predictive power as attitudes themselves. Thus, while these attitude questions are helpful in explaining fixed differences between individuals, changes in reported attitudes are not helpful in explaining changes in outcomes.

In row 9 we investigate whether answers to reservation-wage questions are correlated with future income. Are individuals who report a higher reservation wage today likely to earn more in the future? We see a very strong relationship between reservation wage and future income, even after controlling for the individual's education, sex, and race. This holds true even if we add controls for family background [column (ii)] or family background and current income [column (iii)]. However, changes in reported reservation wages do not help predict changes in income [column (iv)]. In summary, answers to reservation-wage questions do appear to capture some unobserved individual characteristics and might be worth including when trying to predict individual income. Changes in reported reservation wages, however, provide no information about changes in income.

Finally, in the last row, we ask whether answers to job-satisfaction questions help predict future job turnover. Again, we find that people's self-reported satisfaction with their job "as a whole" is a strong predictor of the probability of changing jobs in the future.[3]

IV. Conclusion

Four main messages emerge from this discussion. First, a large experimental literature by and large supports economists' skepticism of subjective questions. Second, put in an econometric framework, these findings cast serious doubts on attempts to use subjective data as dependent variables, because the measurement error appears to correlate with a large set of characteristics and behaviors. For example, a drop in reported racism over time may simply reflect an increased reluctance to report racism. Since many of the interesting applications would likely use these data as dependent variables, this is a rather pessimistic conclusion. Third, and on a brighter note, these data may be useful as explanatory variables. One must, however, take care in interpreting the results since the findings may not be causal. Finally, our empirical work suggests that subjective variables are useful in practice for explaining differences in behavior

[3] In this case, we are not able to study a fixed-effect model, as the job-satisfaction question was only asked in the second and third follow-up of the data.

across individuals. Changes in answers to these questions, however, do not appear useful in explaining changes in behavior.

REFERENCES

Bertrand, Marianne and Mullainathan, Sendhil. "Do People Mean What They Say? Implications For Subjective Survey Data." Mimeo, University of Chicago, 2000.

Sudman, Seymour; Bradburn, Norman M. and Schwarz, Norbert. *Thinking about questions: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass, 1996.

Tanur, Judith M. *Questions about questions: Inquiries into the cognitive bases of surveys*. New York: Russell Sage, 1992.