

Decisions Made About Implicit Attitude Measurement in the 2008 American National Election Studies

Jon A. Krosnick and Arthur Lupia

September 2008

Introduction

One of our main goals in planning the 2008 ANES Time Series Study and the 2008-09 Panel Study was to provide the widest feasible array of reliable measures of constructs for analysts who wish to understand vote choice and turnout.

One factor that may sometimes complicate such measurement is social desirability response bias. Researchers have long suspected that some survey respondents may choose to misreport their attitudes and beliefs and behaviors in order to appear more admirable to an interviewer. Such behavior is an instance of self-presentational social desirability bias. Researchers also suspect that some respondents may engage in self-deception social desirability response bias, whereby people's unconscious leads them to misperceive themselves and to describe themselves inaccurately in surveys.

Concerns about social desirability have led social psychologists to develop new ways to measure attitudes, beliefs, and behaviors so as to get around distortion due to social desirability response bias. One such approach is the measurement of implicit attitudes. One well-known means of acquiring such information is Project Implicit's Implicit Attitude Test (IAT). Another increasingly prominent measure is Keith Payne's Affect Misattribution Procedure (AMP)

Over a period of more than two years, we evaluated existing research on these procedures, commissioned our own studies, sought the advice of experts on racial attitudes, and conducted our own evaluations. Along the way, we have learned a lot about the challenges of implementing such procedures. This report describes some of what we learned and, ultimately, factors that were decisive in the conclusion we reached.

In the end, we decided to include both the IAT and the AMP in our surveys, and we explain below why we made these decisions and how the implementation will occur.

Construct Premises (*P=premise, C=conclusion*)

P. After considering a wide array of approaches for reducing social desirability pressures, we settled on three:

- 1) In the face-to-face interviews conducted for our time series study, we will employ ACASI (Audio Computer-Assisted Self Interviewing). This involves:
 - a. the interviewer turning the laptop around to face the respondent,

- b. the respondent putting on headphones,
- c. the computer playing an audio recording that reads the questions and answer choices aloud to respondents,
- d. and the respondent answering the questions privately by typing responses on the computer keyboard.

Accumulating evidence indicates that this measurement approach reduces social desirability pressures and elicits more honest and less flattering self-reports.

- 2) Our panel study is being conducted using the same approach over the Internet, whereby respondents answer questions on computers in their own homes without the presence of an interviewer. Accumulating evidence indicates that this measurement approach also reduces social desirability pressures and elicits more honest and less flattering self-reports.
- 3) Measurement of implicit attitudes on the time series and panel studies.

With regard to the third technique, we considered the full array of implicit measurement techniques in the literature and quickly narrowed our focus to the IAT and the AMP. Indeed, we were aware of such techniques before we wrote the grant for this election cycle. This methodology's emergence in scientific debates about attitude measurement provided us with a basis for believing that value could come from attempting to integrate such a measurement approach into the ANES.

P. Note. We initially corresponded with authors of the IAT and broadened our thinking to include the AMP in early December 2007 after we became concerned about possible challenges in administering the IAT on our time series and panel studies. Since that time, we have evaluated both measures using common criteria with an eye towards including either, both, or neither.

P. The IAT provides a measure of respondents' preference between two groups (or two other objects or object categories). IAT scores are often interpreted as indicating that a person prefers one group over another.

P. The AMP measures attitudes toward a single object or object category. Such attitude measures can be integrated into a measure of preference by comparing a person's attitudes toward two object categories.

P. We believe that measures such as the IAT and the AMP may have long term relevance. If scholars find such measures to be useful and valid, the ANES should consider including them in many future studies to allow scholars to document and analyze change over time in the nation's implicit attitudes.

P. We believe that all else equal, the ANES user community would be better served by being given valid measures of attitudes toward whites and attitudes toward blacks, from which preferences can then be derived. Providing only a measure of preference between the two groups raises the question of why, if one wants to tap attitudes toward blacks,

whites should necessarily be the comparison group. Perhaps it would be informative to also measure preferences between blacks and Hispanics or between blacks and Asians. We saw no a priori theoretical grounds to justify measuring only preferences between blacks and whites. But we knew we did not have the time available on our questionnaires to measure preferences between blacks and a variety of other racial groups. For this reason, the AMP's ability to generate measures of attitudes toward a single group was appealing.

P. The AMP is a relatively new measure that has not been used nearly as much in past scholarship as has the IAT, which has received wide attention for many years. Many experts whom we consulted believe that the AMP is the state of the art in this area.

P. One complication with measurement of implicit attitudes in the ANES involves Barack Obama as a candidate. In order to predict vote choices in this election, it may be of value to measure implicit attitudes toward blacks as a group. But it may also be of value to measure implicit attitudes toward Barack Obama, because he may be viewed as atypical of blacks by some respondents, due to subtyping. We therefore sought a measure that could be used to measure both sets of implicit attitudes.

Implementation Premises– bases for our conclusions, including findings from research and evaluations

P. Almost all of the data accumulated on the properties of the IAT and the AMP come from two sets of research participants: college students required to complete the measures to fulfill a course requirement, and members of the general public who know the purpose of the IAT and chose to visit the Project Implicit website to learn about themselves.

P. In order to feel comfortable including the IAT or the AMP on an ANES survey, we felt we needed evidence that (1) the vast majority of ANES respondents would complete the measure without complaining about monotony or the length of administration time, and (2) the vast majority of respondents would find the procedure non-offensive. We believed that existing evidence from college students and Project Implicit website visitors was insufficiently informative about how a general public sample, who are not told the purpose of the test, would react to it.

P. After we initiated conversations with the Project Implicit directors, and after several communicative iterations, we all came to a common understanding of the limited value of existing information for our purposes.

C: We determined that we needed to collect new information.

P. Respondents participating in the ANES Panel Study regularly complete questionnaires via computers. Therefore, these individuals are familiar with how to use a computer, which equips them to implement the IAT or the AMP with relative ease. However, many members of the representative sample of Americans who serve as respondents in the Time Series face-to-face interviews in their homes will not be familiar with computers

and will therefore find use of a computer to be challenging. For example, learning to use a mouse for the first time would be required of many respondents, and this would take time we could not afford to spend during the Time Series interviews.

P. A related challenge is that our nationally representative high response rate face-to-face sample tends to be older than that of the Internet panel or volunteer samples of people who choose to participate in internet surveys in exchange for money. Therefore, the experiences of the ANES in attempting to implement the IAT or the AMP would likely be more difficult and time consuming than the experiences of other data collections. We regarded timing and performance estimates from internet-based populations as best case scenarios for the face-to-face survey and interpreted all such results with a substantial probability that face-to-face implementation will prove more difficult and time-consuming.

P. Because of the risks inherent in running either of these data collection mechanisms on an ANES survey, risks such as respondent frustration and increased likelihood of attrition, we decided early on to restrict the placement of such mechanisms – should we run one at all – to locations in the interview schedule where they could do minimal damage to the rest of the planned data collection. For the Time Series, this meant considering these measures only if they would be placed at the very end of the post-election interview, in order not to risk causing some respondents to decline to participate in the post-election interview if they did not like doing the IAT or the AMP during the pre-election interview.

P. Also affecting our considerations for the ANES Time Series was a time constraint. Both the IAT and the AMP would require the use of ACASI software. ACASI software allows the respondent to enter responses directly on to a laptop rather than stating verbal answers to a trained interviewer and then having the interviewer record the response on the laptop. We are contractually obligated to have no more than 10 minutes of ACASI-based questions on the post election wave of the Time Series, and some of those ten minutes will be occupied by questions on other topics. That means that we could not implement an implicit attitude measurement procedure that might last longer than the remaining amount of time within the 10 minutes for any respondents.

P. The ACASI time constraint meant also that the IAT or the AMP was competing for those ten minutes with other self-reports whose validity could be increased by allowing respondents to answer the questions confidentially. This competition was especially stiff, because the involvement of Barack Obama in the race for President meant that we needed to implement various explicit racial attitude measures, which we believed would be most effectively administered via ACASI.

P. The more general opportunity cost of including an implicit attitude measure on the Time Series is the ability to include other questions. If one of these measurement tools requires five minutes to administer on average, then – given our experience of being able to ask 3-4 ordinary self-report questions per minute – we would lose the ability to ask 15-20 other explicit questions. We therefore had to decide whether the value of an implicit

attitude measure to the user community would outweigh the value of asking 15-20 questions that would be eliminated.

P. The IAT has multiple versions. The most famous version takes much more time to administer than the *brief IAT*. We determined that there was no way we could include the regular IAT in our surveys because of administration time. Hence, we worked with Project Implicit to evaluate properties of and implementation challenges associated with the “brief IAT” (which henceforth, we will simply refer to as the IAT). The brief IAT evaluated in Sriram and Greenwald (2007) required 40 trials. It was evaluated on student samples of psychology class enrollees at the University of Washington.

P. We began with no direct estimate of how long it would take to implement the IAT or the AMP on the Time Series.

P. We began with no reliable estimate of the amount of time it would take to transition to and explain either mechanism in a face-to-face survey. At best, we have aggregate timings from non face-to-face platforms. It is likely that the “fixed cost” (in terms of time) of explaining the method is quite different than the marginal cost of completing additional trials.

P. All ANES PIs and lead staff have taken both the IAT and the AMP. All of us found the former frustrating at one time or another. For Jon Krosnick, this frustration occurred in public when he attempted to demonstrate the IAT at a conference at the Gallup Organization’s world headquarters in Washington, DC, that we held with many of the nation’s leading news media pollsters. Many in attendance remarked that they did not believe that we could get many members of a nationally representative sample to complete the exercise.

P. Not wanting to rely strictly on student samples or on samples of people who chose to visit the Project Implicit website, we commissioned additional data collections for each measure in order to observe and evaluate the performance of these measures with samples that are more similar to ANES respondent samples.

P. We worked with Project Implicit to build a “partisan brief IAT” to measure preferences for Republicans and Democrats, which took Matt DeBell 7 minutes to complete in December, and he felt the experience was frustratingly tedious. An IAT focusing on race would take less time. An equivalent report from Matt about the AMP time estimate was 1-2 minutes. Matt is faster and more experienced than our typical respondent would be.

P. On May 2, 2008, after implementing the IAT with a large sample of American adults who volunteer to do internet surveys for money (provided to us by SSI) but did not know what the IAT is, we discovered the following (as described by Matt DeBell):

“The mid-interview termination rate on the IAT is looking alarmingly high. ... SSI reported that a typical rate is 15%. We are losing 40% on the instrument

from start to finish. Some of those could be from people failing to continue after the debriefing page at the end, and we'll figure out those details soon. But if we look only at the ones who got to the beginning of the IAT task and didn't continue to the debrief page at the end, we are still losing 28%.

The high breakoff rate implies Rs don't like the task. (It could also be due to technical problems.) Unless the picture changes, this would make me hesitate to include the IAT on the Panel Study before the very end of the May 2009 questionnaire.”

P. We then contacted the SSI respondents who did not complete the measure to ask why they did not. They expressed a variety of different types of frustration. Some frustration was clearly due to problems relating to download speeds. This would continue to be an issue on the Panel Study due to the variety of web platforms by which our panelists connect to KN. One of us noted:

“On the panel, we are reliant on whether the respondent's computer has a fast internet connection and can run flash - we're learning that this is often not the case. So there's the technical problem that will limit the size of the sample that may be able to complete the measure, and we saw that some people think they've been frozen or kicked out by the software, so we would need to put the IAT at the end of a panel wave to prevent other damage.”

P. Some of us were concerned that putting a frustrating measure at the end of a wave would increase the attrition risk because – as the last item on the wave – it would be more likely to stick in memory as representative of the experience of completing the survey than other content (here, we were referring to Kahneman's work on the peak and the end of remembered experiences influencing judgments of pleasantness).

P. Based on all this information, we decided to test the AMP with Knowledge Networks respondents.

P. On June 10, 2008, we learned that running the IAT on the Time Series Survey would require purchasing licenses for special software that would have to be installed on each laptop and reconciled with the existing interview software. We did not realize that there would be additional costs incurred by administering the Time Series Surveys, and we had no funds budgeted to cover such costs. The cost issue came up when Project Implicit personnel realized that we could not collect data for the Time Series on the Project Implicit website, because the interviewers' laptops would not be connected to the Internet during the interviews.

P. Our rough estimate of this expense was about \$8,000 plus additional costs that would come from attempting to reconcile the programs. None of the other questions in the questionnaire would benefit from this purchase. The AMP had no similar expense associated with its implementation, and its software was simpler.

P. Knowledge Networks told us that they could program an IAT from scratch more easily than they could incorporate PI's software into their system, and that they did not anticipate difficulty doing that from-scratch programming.

P. On June 13, 2008, Project Implicit and RTI reported outcomes from a direct meeting. Darrell Donakowski conveyed the following information about that interchange:

“When [DELETED] initially asked about the timing of the IAT, Sriram said it was between 5-10 minutes. I mentioned that I thought there was a short IAT that we would use, and Sriram said that it would still be between 5 and 7 minutes. He mentioned that time is needed for the interviewer to instruct the respondent and for the respondent to go through two test cycles.”

P. On June 24, 2008, we received information on the outcome of the testing of the AMP. The psychometrics looked good, and evidence of validity was convincing. Furthermore, implementation of the AMP was less challenging for respondents and certainly no worse than for the IAT.

Conclusions

C. We concluded that there is a substantially greater risk of respondent frustration and attrition associated with the IAT than with the AMP.

C. The evidence we acquired suggests that the AMP could be administered in much less time than even the Brief IAT. We also concluded that the AMP requires less respondent training – a critical factor for the Time Series, where the opportunity cost of time is very high and the sample is likely to “trend older” and, hence, have fewer computer skills than samples on which we and others have evaluated AMP and IAT.

C. We believe that the use of multiple proprietary software programs in order to implement the IAT produces the prospect of serious production risks. The AMP software is more flexible and portable.

C. There is no precedent that we know of for having the ANES pay an extra expense for the purpose of administering a single question. The precedent set by deciding to pay such a cost, and the perceptions of favoritism that such treatment would likely engender among the community of scholars who would like ANES to direct scarce resources to their research agendas, would take us into unknown territory with respect to the project's perceived legitimacy. This factor is not determinative in our decision, but its relevance is non-zero.

C. *Time Series.* Taking all of these factors into account, we concluded that there is value to the user community of placing one of these measures on the Time Series. Given that the measures have similar psychometric properties and comparable breakoff rates, and given our judgment that the AMP's measure can be implemented in far less time than

even the brief IAT, we decided to include the AMP at the end of the post election wave of the Time Series.

C. Time Series. We decided to include on the Time Series a version of the AMP that shows photos of blacks and whites. We debated whether to run an AMP measuring attitudes toward Obama and McCain instead. However, two rationales led us to focus on blacks and whites. First, the latter measure may have value not just in 2008 but also when compared to similar measures administered in future years. Interest in other black candidates may arise. If McCain defeats Obama, Obama may not be a central figure in future election cycles. Second, many research agendas can benefit from an improved measure of attitudes towards blacks and whites. By the end of the election season, many respondents may not think about Obama as typical of blacks. As a result, an Obama-focused AMP may have limited usefulness then a more general measure of attitudes toward blacks.

C. Panel Study. Absent additional information about the IAT or the AMP that presents starkly different results about their likely performance with a representative sample, we also decided to include the AMP measuring attitudes toward blacks and whites on the Panel Study. In September 2008, half of the respondents will get the AMP focused on blacks and whites, and the other half will get the AMP focused on Obama and McCain. In October 2008, the opposite will occur.

C. Panel Study IAT. Moreover, we have decided to place the IAT measuring preferences between blacks and whites on a non-political post election wave (November or later). This can be done on the panel because the opportunity cost of time is less. The Time Series has built into it a CORE set of questions, which takes about 60% of the entire survey to administer, as well as modules committed to other purposes such as the CSES, DHS, and Bonus Minutes programs. Its long and visible history also creates extensive demand for issue and topical coverage. The Panel Study is new. It has neither the obligations nor the expectations built into it. As a result, we can devote more time to items that are high-risk high-reward. The fact that we can include this measure after we have collected all of the needed Fall 08 electoral information means that the cost to the instrument of increased respondent frustration or attrition is less than would be the case on the Time Series or on earlier waves of the Panel Study.

Coda

We are very grateful to everyone at Project Implicit and to Keith Payne for working with us through this process. They have been patient and accommodating, and their efforts have done a great deal to advance scientific study of political attitudes in the American National Election Studies.