Political Attitudes and Response Bias in Semi-Democratic Regimes
A Survey Experiment Comparing the List Experiment and Randomized Response in Tanzania[*]

Keith R. Weghorst[†]

August 2015

[†]Post-doctoral research fellow, Department of Political Science, Vanderbilt University **Address**: PMB0505, 230 Appleton Place, Nashville, TN, USA. 37203 **Tel:** +1.847.644.2890 **Email**: keith.r.weghorst@vanderbilt.edu

Abstract: In semi-democratic regimes where a large degree of self-censoring exists, standard survey questions give us little traction for eliciting responses on sensitive, yet extremely compelling political issues. Randomized Response Technique (RRT) and list experiments have emerged as means of improving the quality of survey data on politically sensitive issues. This paper employs both to investigate the degree of measurement error in eliciting responses to conventional survey questions about support for opposition violence against the incumbent party, the independence of police and military, and the fairness of electoral competition in Tanzania. Results from a survey experiment conducted in May-June 2009 with University students show that using RRT and list experiments produce remarkably different results compared to conventional question formats. These findings suggest that more accurate measurement of citizens' sensitive viewpoints is attainable through list experiments, opening up an avenue of research from which scholars have previously been deterred in anticipation of low-quality data. This paper presents, to my knowledge, the first attempt to provide directly comparable results of a list experiment, randomized response technique, and direct question formats about politically sensitive topics in a semi-democratic state.

1 Introduction

Voicing contentious opinions poses risks to individuals and questions about them consequently tend to yield poor quality public opinion data. This includes expressing opinions that are socially undesirable—holding prejudiced views against other racial or religious groups—or actions that are otherwise unacceptable—drug abuse, theft, and other illicit activities. In politics, voters routinely report having participated in previous elections regardless of whether or not they cast ballots;[1] and survey subjects say they support the candidacy of a female or ethnic minority for head of state even if they are actually angered by such campaigns. Fear of social sanction from peers drives them to give what they perceive as the "right" answer to a survey question. This bias can yield very poor data, marred either by high (and systematic) rates of item non-response or inaccurate responses that do not reflect the "true" opinion of the survey respondent.

We have little insight regarding how problematic sensitivity-induced response bias is in semi-democratic regimes and whether alternative question formats might attenuate its effects. We do know that in some semi-democratic regimes, respondents hide their "true" attitudes with respect to political participation rights and vote-buying practices (Corstange 2009, Gonzales-Ocantos et al. 2011).[2] For parsing out the dynamics of opposition politics in uncompetitive electoral regimes, these issues are critical and have decreased willingness amongst researchers to pursue topics of substantial importance in both the academic and policy world in anticipation of poor measurement. Support for political opposition in such regimes poses more substantial risks than social sanction: it can actually mean the difference between life and death for many citizens. Where respondents believe that surveys are conducted by government agencies—as do 59.9% of Tanzanian subjects in Afrobarometer's 2008 survey—the accuracy of estimates of public opinion of even relatively innocuous issues like trust in state institutions and incumbent performance might be seriously limited by response effects.[3] I present results from a survey experiment conducted at the University of Dar es Salaam, Tanzania in 2009 testing the relative performance of two alternative question formats designed for parsing out sensitive political attitudes. This paper joins a growing literature on the application of list experiments

---

[1]See Abramson and Claggett (1984), Bernstein et al. (2000), Cassel (2003), Duff et al. (2007), Holbrook and Krosnick (2010b).

[2]By semi-democratic regimes, I mean regimes that "play the game of multiparty elections...yet they violate the liberal-democratic principles of freedom and fairness so profoundly and systematically as to render elections instruments of authoritarian rule" (Schedler 2006: 3). Conceptually, this resembles what has elsewhere been called "competitive authoritarianism" (Levitsky and Way 2010).

[3]Afrobarometer Survey Tanzania, Round 4, Question 100: "Who do you think sent us to do this interview?" 43.54% of respondents felt the survey was conducted but the general umbrella of "government," while the remaining 16.4% named a specific governmental figure or agency.

(the item count technique) for studying a variety of political attitudes in different cultural contexts.

2   Collecting Sensitive Survey Data

Research on survey methodology has illustrated that certain question formats perform better than others in eliciting sensitive information from respondents. Because conventional questions about sensitive topics pose heightened risks for subjects, less invasive measurement strategies that indirectly ask about sensitive topics might be better suited for studying things like the legitimacy of opposition violence. Doing so attenuates concerns of subjects over secrecy by masking individual responses and instead producing data suitable for group level analysis. There are two frequently used techniques for questioning respondents about heated political issues through less obtrusive means. The first employs randomized assignment to innocuous and sensitive questions and the second has respondents aggregate responses from lists of political attitudes where only some lists contain the sensitive item.

A first approach is known as Randomized Response Technique. First, respondents are randomly assigned to answer one of two questions. One of the questions is about the sensitive item of interest and the other an innocuous one which we can expect respondents to answer truthfully. Both questions share the same response options—typically "Yes" and "No." The researcher can only observe the total proportion of "Yes" responses and usually designs a non-sensitive question for which an estimate of the total "Yes" responses is known. For example, if the non-sensitive component instructed respondents to roll a die and answer the question "Did you roll a 5?" we know that respondents will answer that questions in the affirmative about 17% of the time.[4] Knowing the overall proportion of "Yes" responses and the expected proportion of "Yes" responses for the non-sensitive item, we can estimate the proportion of respondents who answered the sensitive question "Yes."[5] A disadvantage of the RRT is its relative inefficiency because assignment to answer either the sensitive or non-sensitive question is induced by a probabilistic event like flipping a coin. Using probabilities to estimate the proportion of responses to the non-sensitive question also increases the imprecision of the measure. One can overcome the first issue by knowing the actual proportion of respondents assigned to the sensitive and non-sensitive question, an option available when using self-administered questionnaires. Regarding the second issue, one can take advantage of a closed environment like a classroom where the value of the innocuous question is known, like the number of females in the room. RRT has

---

[4]The "die roll" technique was introduced in Warner (1965)

[5]Please find the formula in the appendix.

been used to shed light on a number of issues in the developed world,[6] can be used with illiterate populations

in the developing world,[7] and now has means of conducting more sophisticated analysis of data collected

through the technique.[8]

The other less obtrusive question format is commonly called a "list experiment" or the "item count tech-

nique" because it asks respondents to indicate the number of items on a list of behaviors they engage in

(Rayburn et al. 2003, Tsuchiya et al. 2007, Wimbush and Dalton 1997), attitudes they agree with or oppose

(Kuklinski et al. 1997a), or things that anger or upset them (Kane et al. 2004, Martinez and Craig 2010,

Sniderman and Carmines 1997, Streb et al. 2008). By giving some respondents instruments with a list that

contains a set of items which are by design disagreeable but not controversial, and randomly assigning other

respondents to a list with an additional sensitive item, one can estimate levels of support for the sensitive

item within the group.[9] List experiments have proven effective for identifying socially undesirable attitudes,

including racial prejudice (Kuklinski et al. 1997a,b), anger over a female presidential candidate (Streb et al.

2008), and biases towards certain religious groups (Kane et al. 2004). Applications of the list experiment

have expanded dramatically in the developing world as well, with recent applications to vote-buying in Latin

America[10] and electoral rights in the Middle East.[11]

Both the list experiment and randomized response are theorized to better capture "true" public opinion,

yet we rarely have evidence of *how much better* the alternative formats perform. In some cases, the relative

power of these question formats over conventional designs is established indirectly, with authors offering the

list experiment as a more authoritative voice amongst studies within a research program that with conflicting

empirical findings. (Kane et al. 2004, Kuklinski et al. 1997a,b, Sniderman and Carmines 1997). Others have

employed designs that permit within study comparison of responses from conventional formats and the list

technique.[12]

Yet, the use of list experiments and randomized response imposes an "unavoidable precision cost."

---

[6]For an overview, see Chaudhuri and Mukerjee (1988) and Holbrook and Krosnick (2010a).

[7]Soloman (2007) utilized RRT for studying illegal park resource usage with illiterate subjects in Uganda.

[8]Gingerich (2010)

[9]Items on the list must be designed to reduce the possibility that respondents agree with all of the items or none of them. Both "ceiling" and "baseline effects" would undermine the goal of the list experiment; individuals who agree with all of the items will identifiably agree with the sensitive one, as will respondents who agree with a sensitive item *and* respondents generally agree with none of the items on the treatment list (Glynn 2009, Kuklinski et al. 1997b).

[10]Gonzales-Ocantos et al. (2011)

[11]Corstange (2009)

[12]Corstange (2009), Gonzales-Ocantos et al. (2011) are of particular note.

(Corstange 2009: 62) Data collected with list experiments and the randomized response technique have been primarily analyzes through univariate, group level comparisons.[13] Moreover, while common sense informs us *why* both these formats would produce better data, rarely are such insights supported by within study comparison of *how much* better the techniques operate in comparison to direct questions.[14] In order to evaluate the benefits of using these alternative question formats instead of conventional ones, we must develop a baseline comparison of support for things like opposition violence *from within a survey sample.* Finally, because RRT and List Experiments both impose additional but different cognitive demands on survey subjects, we look to establish whether or not they "work" with respondents in a developing country and which alternative performs better.

3 Case Selection

3.1 Tanzania at the Macro-Level

Tanzania is a prototypical dominant party system, a common variation of electoral authoritarian regimes.,[15] In Tanzania, elections take place, but in a heated environment where claims of fraud, disenfrachisement, and repression are not uncommon. Like many African countries, shortly after independence the ruling Tanganyikan African National Union (now called Chama Cha Mapinduzi (CCM)) outlawed multi-partyism and the party came to control nearly all political, social, economic life throughout the country (Barkan 1994, Hyden 1980, Mmuya and Chaligha 1994, Msekwa 2006). Tanzania's "top-down democratization" (Hyden 1999) in the early 1990s reintroduced multiparty competition, but opposition parties managed only to attain 10.5% of seats in the 1995 elections, 12.6% in 2000, 14.9% in 2005, and about 25% in 2010 (Msekwa 2006, TEMCO 2006). The opposition is partly hindered by institutional rules that mechanically disadvantage political challengers through disproportionality.[16] Challengers often point to abuse of incumbency, fraud, threats, and violence as drivers of their lack electoral success because voters fear to support them. This suggests that we have good reason to expect data on political attitudes in Tanzania might be influenced by

---

[13]Corstange (2009), Glynn (2009), Imai (2011) have recently developed estimation procedures that permit multi-variate and individual level analysis.

[14]Notable exception are Coutts and Jann (2008), Gonzales-Ocantos et al. (2011).

[15]I use the term "dominant party system" to describe a regime in which a party has held the executive office and a legislative majority for an extended period of time. Other similar terms include "single party dominant system" and "hegemonic party system." For a review of different criteria of party "dominance," see Bogaards (2004).

[16]Tanzania is a presidential system with 239 legislative seats elected in single-member districts by plurality rules. An additional 102 seats are allocated to females of parliamentary parties proportionally, five are guaranteed for Zanzibari representatives, and up to ten may be appointed by the President.

sensitivity biases.

3.2 Potential Response Bias in Tanzanian Survey Data

As we might expect, public opinion data on political views in Tanzania reflects this sensitive political context. This study is admittedly limited in how well it can dialogue with nationwide public opinion data collected in Tanzania. Nonetheless, a brief review of a study conducted in 2001, one year after the country's second multi-party elections, helps ground concerns about response bias in expressing political attitudes in the context of Tanzania. In this first round Afrobarometer survey,[17] Tanzanian respondents repeatedly reported the highest levels of trust in state institutions across all Afrobarometer countries, including the Presidency (91% of respondents trust the president "somewhat" or "a lot"), the judiciary (72%), the military (95%), and the electoral commission (82%). Moreover, 88% of citizens claimed to be satisfied with the performance of President Benjamin Mkapa in the first year of his second term, while only a little over half of the respondents thought that the government had successfully reduced corruption during the same period. Given that anti-corruption was the major campaign platform on which Mkapa ran, such a result suggests something is amiss.[18] Most surprisingly, the 50% of respondents that felt the 2000 elections were free and fair was one of the highest portions of respondents the round of Afrobarometer surveys. This was in spite publicized disenfranchisement, rigging, and polling day violence that led the elections to be labeled as "irregular" and "less than free and fair" (TEMCO 2006). Viewed in the light of response bias, these seemingly conflicting responses resulted neither from Tanzanians being "uncritical citizens" nor "patient trustees" (Chaligha et al. 2002). It is far more plausible that in semi-democratic context, one simply cannot collect reliable data about heated political topics like incumbent performance and opposition support using conventional survey formats. The consequence is that systematic survey biases mask the shape of politics on the ground.[19]

---

[17] I use 2001 data here to suggest an alternative interpretation of Chaligha et al. (2002)'s finding of inconsistent viewpoints of Tanzanians in the 2001 study. Although not in such precise terms, the authors identify response instability and a lack of constraint amongst Tanzanians (See: Converse 1964) and argue occurs because of unique political legacies that shape these attitudes. I would instead attribute to a political legacy of semi-democratic governance that induces measurement error (Achen 1975, Althaus 1998). Later in the paper, I use more recent Afrobarometer data.

[18] See Mkapa (2003)

[19] On issues of systematic non-response and response effects, see Acock (2005), King et al. (2001).

4   Research Design

4.1   Sampling Frame and Survey Administration

From May 8-June 1, 2009, I administered a survey in Tanzania that assessed the presence of sensitivity bias when eliciting political attitudes with conventional question formats, the list experiment design, and the randomized response technique. The data was collected using a convenience sample of sophomore, junior, and senior political science majors at the University of Dar es Salaam. As a research associate of the University, I was granted access to public administration and political science lecture courses, where the surveys were conducted. Respondents completed self-administered questionnaires in a lecture hall setting during a normally scheduled class period. Experiments conducted with students in a lab setting have some disadvantages in terms of the ability to generalize the findings of a study to other non-student populations.[20] The aim of this project, however, was to test the function of the question formats. A lab setting where there are high levels of control over the treatment administration and the experimental environment is ideal for this goal. Additionally, because completing a questionnaire in a classroom setting is a fairly common behavior for a student to engage in, this design also increases the ecological validity of the study. We know that students in sub-Saharan Africa tend to be more outspoken than other population subgroups and that self-administered questionnaires tend to reduce sensitivity biases. This gives us good reason to expect that response bias would be lower for this subgroup and instrument format and thus, presents a "hard test" of question format performance.

After the distribution of the instruments, the classroom was given a brief introduction to the project. This included reading the verbal consent script, a reminder to indicate item counts on the list experiments (as opposed to individual items) and where on the instrument the information on group assignment for the Randomized Response Technique could be found.[21] Respondents were then instructed to complete the questionnaire and informed they would be collected at the end of the class period. In total, I distributed 430

---

[20]I present my study as a survey experiment because the experimental treatment originates from different assignment of individuals to different survey instruments, but it shares many attributes with lab experiments, including the controlled environment. For more on the distinction, see Morton and Williams (2010).

[21]Both debrief sessions and two waves of a subsequent study in Dar es Salaam have shown that both careful instruction to aggregate (not indicate) items and a visual reference of the list is essential for the list experiment to function properly.

self-administered questionnaires and collected a total of 261 completed surveys.[22][23]

4.2   Instrument Design

The survey focused on three sensitive items which varied in their controversial nature: support for the use of violence by political opposition, the extent to which the ruling party sabotages attempts of the opposition to compete in elections, and the independence of the state repressive apparatus. I selected these topics for three key reasons. First of all, they represent some "likely candidates" for difficulties in collecting accurate data on politics in semi-democratic states and issues central to establishing the democratic nature of political competition in these regimes. Secondly, they vary across each item in terms of sensitivity: offering support for violence against a regime is far more controversial than believing that an incumbent party might misuse the police.[24] Third, these topics have been studied in other public opinion studies— albeit with different wording—offering at least some insight on what settings and topics might benefit most from list experiments and randomized response. The instruments also included a few basic political attitude questions—such as partisanship and interest in politics—but were kept to minimum because the ambition of the project was not to understand student political attitudes.

The experiment featured three sensitive questions and three different question formats. I randomly assigned respondents into three different groups, using a split thirds method. The idea was that a respondent would receive each of the three questions, one designed as a list experiment, one designed as a randomize response format, and one as a direct question. Across the three groups, the technique used to pose a particular question differed. Table 1 provides an illustration of this design and the distribution of subjects across the groups. If a respondent was assigned to the first group, she was asked about the use of opposition violence directly, about incumbents sabotaging opposition embedded in a list with four non-sensitive items, and incumbent control over the police and military with the randomized response technique. In the second group, the instrument used the randomized response technique to ask about incumbent sabotage, asked

---

[22]According to AAPOR standard definitions, the study had a response rate of 60.7%. All students were given a questionnaire and were informed that if they did not wish to participate, to leave the questionnaire blank and to return it as such.

[23]Most response effects relate to surveys conducted by an enumerator. Thus, if any bias comes out of using self-administered questionnaires it would be against response effects due to political sensitivity.

[24]While this is intuitive, consider that in 2006, 11% of respondents for a 2006 Afrobarometer poll agreed that "it is sometimes necessary to use violence in support of a just [political] cause," while about 50% had "a lot" or "a very great deal" of trust in the police and military.

about incumbent control over the police and military directly, and opposition violence along with four items

other items, different from ones used for the first group. The third group received the incumbent sabotage

question directly, a randomized response format for the opposition violence question, and a third unique list

of non-sensitive attitudes embedded with item on control over the police and military. The use of the list

experiment for three different questions increases the complexity of the design slightly because each of the

list experiment "treatments" containing the sensitive item requires a unique list of items and a control group

for each the three lists that does not contain the sensitive item. For the group of respondents who received the

list experiment treatment for the opposition violence question, then, this group was assigned to the control

group for the other two list experiment questions and received these lists without the sensitive incumbent

sabotage and military/police items. For every sensitive question, there is one treatment group that received

the list experiment with the sensitive item and the two other groups who received the control lists. This

means that the control group for each list experiment is roughly twice the size of the treatment group. The

randomized response technique also utilizes the logic of treatment and control groups, but it occurs within

assignment to the randomized response group so that only respondents who, for example, received RRT for

the opposition violence question can be in the treatment or control groups of the RRT estimate of support

for opposition violence.

INSERT TABLE 1 ABOUT HERE

4.3   Question and Content Design

Direct Question Format I asked each of the three direct questions using a prompt, followed by paired

response options. Questions using this format appeared as follows:

> **Opposition Violence**: "Some people believe that when the opposition is pushed too far in politics, it may be
> necessary for them to use violence. Others think that violence is never acceptable in politics. Which of these
> statements do you agree with more? How strongly do you agree?"
>
> Statement 1: Under some circumstances, the political opposition must use violence.
>
> Statement 2: The political opposition should never use violence in any circumstances.

Incumbent Sabotage: Which of these statements do you think better reflects politics in Tanzania? How strongly do you agree with the statement?

Statement 1: The reason opposition parties are unsuccessful in elections is because the ruling party sabotages their attempts to compete.

Statement 2: The reason opposition parties are unsuccessful in elections is because their policies are not as good as Chama cha Mapinduzi's.

Incumbent Control of Police/Military:Which of these statements do you think better reflects politics in Tanzania? How strongly do you agree with the statement?

Statement 1: The military and police serve Tanzanians fairly.

Statement 2: The military and police primarily serve the interests of the ruling party.

List Experiment Technique

Respondents were asked about one of the three sensitive items by a list experiment. Each of the three list experiments used four unique control items that are somewhat controversial and for which Tanzanians often have strong opinions. No control item, however, was politically sensitive like the treatment item. As an example, I provide one list below and the rest in the appendix.

Opposition Violence

"Below, I provide a list of different political attitudes that Tanzanians sometimes agree with:

· Students should contribute some of their stipend to develop the nation.
· Julius Nyerere set us on the right path forward.
· I try to access the news when possible by radio, newspaper, or television.
· The government should reduce the costs of fuel by raising taxes.
· Under some circumstances, the political opposition must use violence.

How many of the items do you agree with? Do not indicate which items, just how many of the five items you agree with. Please circle one of the numbers below.

In order to estimate support for a sensitive item we take the difference between the mean number of items selected by the control group and the treatment group. Given that the treatment is randomly distributed, we expect the mean number of agreed non-sensitive items to be the same for the control and treatment groups. The difference of means between the two groups tells us the direct impact of the adding the treatment item

to the list. A difference of 0 between the two group means indicates 0% of respondents in the treatment group agree with the sensitive attitude, while a difference of 1 would indicate 100% agree with the sensitive item.

Randomized Response Technique

From discussion of Randomized Response Technique above we know that its implementation requires (1) randomly assigning individuals to control and treatment groups and (2) a non-sensitive "Yes/No" control question for which we can estimate the proportion of "Yes" responses. To randomize assignment, I stapled a small sheet of paper on the inside of cover of each paper questionnaire with an image recognizable to Tanzanian students. Respondents were given instructions that upon completion of the survey, they were to remove the picture that had been stapled to the survey. Respondents who received a questionnaire with a picture of a cow were assigned to answer the sensitive question. Assignment to the control group was denoted by a picture of a chicken. Individuals assigned to the control group responded to the question, "Are you female?" By recording the proportion of female students present when surveys were distributed (37%), I can establish a baseline estimate of "yes" responses and then measure deviations from this value. This approach differs slightly from most uses of RRT because the survey experiment is conducted in a closed environment. This permits the use of a known value rather than an expected probability and reduces much of the noise surrounding estimates.[25]

---

[25] Although response rates were remarkably high for self-administered questionnaires, there remains the possibility that the subset of students who returned the questionnaires differs from the group receiving the surveys. When the surveys were collected, I kept a count of the number of females who returned surveys; the number of females submitting surveys was slightly higher in this informal count (1.4%). More on this later in the paper.

> **Incumbent Sabotage**
>
> *Attached to the survey, you received a piece of paper that has a picture of a chicken or a cow. I do not know which picture you received. If you received a picture of a chicken, please answer Question A. If you received a cow, please answer Question B. Do not indicate the question you answer, only circle "Yes" or "No."*
>
> Question A: Are you female?
>
> Question B: Do you think the opposition parties are unsuccessful in elections is because the ruling party sabotages their attempts to compete?

5   Results

Tables 2-4 present the results of the survey experiment by comparing responses generated by the list experiment, randomized response, and direction question formats.[26] Both the aggregation and randomization procedures yield higher levels of support for the sensitive political issues addressed in the survey.

When asked directly about opposition parties' use of violence, about 41% of subjects agreed that sometimes opposition parties must resort to the use of violence. This number is higher than the Afrobarometer finding that 26.1% of Tanzanians under the age of 30 support violence for a "just cause" in politics, indicating this sample of university students may hold more radical views in comparison to citizens.[27] In comparison to the direct question about whether the opposition must sometimes use violence to the other formats, over twice as many respondents indicated support for this statement through the randomized response technique and the list experiment (See Table 2). Both formats indicate over 80% of respondents feel opposition parties must in some cases resort to violence. Of the three sensitive questions included in the instrument, this item was by far the most likely to produce response effects. The massive gap between responses uncovered by the RRT and list experiment suggests that conventional approaches towards these issues produce biased estimates of public opinion.

---

[26]Note that the list experiment portion of analysis uses Welch's t-tests to compare the means of the control and the treatment groups. I choose this test because five of the six lists are normally distributed (via Shapiro-Wilk tests) and it permits unequal variance between treatment and control groups. I also used a non-parametric Mann-Whitney U test that does not assume a normal distribution. Both robustly confirmed the findings in the paper; I present the t-tests out of convention of research using list experiments.

[27]Based of the Round 3 survey conducted in Tanzania in 2005; the opposition violence question was not included with the most recent round conducted in 2008. The subsample of students under 30 reported levels of support that are similar to what I find for the direct question format (43.5%), but only amount to about 1.5% of the total sample of the Afrobarometer study.

INSERT TABLE 2 ABOUT HERE

We noted that Tanzanians reported extremely high levels of trust in state institutions, including the electoral commission, police and the military. Just like the issue of opposition parties engaging in violent actions, questions about the fairness of political competition and the independence of the state repressive apparatus may produce response effects. This could be because survey subjects try to offer a "right" answer, or one that is safer, more socially desirable, or both. These response effects operate differently than voicing support for a radical behavior. For these items, we would expect answers to questions about political trust to be more generous towards the incumbent regime. Thus, we have good reason to suspect the finding that only 19.1% of respondents under the age of 30 in Afrobarometer's 2008 survey of Tanzanians have little or no trust in the ruling party is at least in part an artifact of question format.[28] Table 3 shows responses to the question I included about beliefs on incumbent sabotage of the opposition.

INSERT TABLE 3 ABOUT HERE

In Table 3, we see that once again the list experiment teases out greater support for sensitive political attitudes. The question about incumbent sabotage is much less sensitive than the first, which asked about the necessity of opposition violence under certain circumstances, as a greater portion of respondents who received the direct question indicated that incumbents interfere with the attempts of Tanzania's opposition parties to compete. It may seem remarkable that only about 10% of the survey subjects thought the ruling party "fights fair." However, given that 9.9% of respondents self-identified as CCM partisans, this accords with what we expect unbiased responses to this question to reflect: those respondents who identify with the incumbent party are likely to think it wins elections legitimately. Surprisingly, we see that the randomized response technique estimates substantially lower levels of belief incumbents sabotage the political opposition. There are many possible explanations for this unanticipated finding but it most likely occurred due to particular problems of administering the RRT portion of the survey, an issue I will discuss later.

An earlier section of the paper showed high levels of trust of the military and police in Tanzania. Indeed,

---

[28]Unfortunately, Afrobarometer did not include an occupation item the fourth round that would let us separate students. For the entire sample of Tanzania, 18.5% of respondents had little or no trust of the incumbent regime.

in round four of the Afrobarometer survey (2008), less than 12.3% of subjects had no trust in the police.[29] Table 4 tells an all-together different story. For each of the question formats, we saw that students of political science at the University of Dar es Salaam feel the police and military are quite preferential in who they serve. About 80% of them indicate that the state's repressive apparatus is discriminatory and that the ruling party uses it to carry out their own agenda.

INSERT TABLE 4 ABOUT HERE

An interesting element of Table 4 is that the gap between responses brought about through the list experiment, RRT, and direction question formats have converged to produce very similar estimates. Figure 4 provides a visual representation of how the differences in estimations from standard questions and ones that use randomization and aggregation techniques monotonically decrease as question sensitivity declines.[30] It was only the question about opposition violence that produced widely different responses for the direct question and the alternative formats. This coheres with what we should expect. If in comparison to opposition violence, the other issues are not very sensitive to University students, then the direct question format for those issues should yield similar results to the list experiment and randomized response. Comparisons to the Afrobarometer data showed that in a representative sample of Tanzanians, only a small portion of respondents support violence that is politically justifiable. About half of respondents have some distrust Tanzania's highly partisan military and police. Stated differently, the list experiment and RRT seem to work best for highly sensitive political issues. More importantly, when it came to issues that were still sensitive, but ones that individuals seemed more willing to talk about openly, each of the three question formats produced very similar results. This means that the list experiment performed at least as well as directly asking a respondent about all three sensitive questions, while the randomized response technique worked well for questions about opposition's use of violence and the independence of the police and military. Using these two alternative approaches towards survey data is a viable approach not only for highly sensitive political topics, but also for ones which are less controversial.

---

[29]The fourth round survey also drops the question about trust in the military. However, the third round shows that less than 20% of respondents have no trust in the military.

[30]Note that the 95% confidence intervals of estimates of respondent agreement with sensitive items displayed in figure 4 are truncated at 100%.

6 Robustness

6.1 Target Population and External Validity

Conducting this survey experiment with university students may in one sense be a generous testing ground simply because interest in engaging complicated questionnaires is likely higher than for the average Tanzanian. Using students for experiments is not uncommon. Indeed, the convention in experimental research conducted in controlled environments tends towards using student subjects.[31] Yet, there remain other potential pitfalls of using convenience samples of political science university students. The core goal of this project has been to test question formats and consequently has placed lower priority on external validity. Nonetheless, the sensitive questions posed resembled ones asked by the Afrobarometer to provide a rough external validity check. The survey was also designed to include responses to the questions posed with conventional formats. Together, this permits a limited comparison of the list experiment and RRT to direct question formats and responses from the direct format of the convenience sample to ones captured with the more representative samples conducted by the Afrobarometer. In comparison to the direct question in both the survey and the subgroup (student under 30) Afrobarometer data, the list experiment and RRT generally showed that the latter question techniques yielded higher levels of support for sensitive political attitudes.

A remaining concern is the external validity of the survey data collected from this sample: what does this information have to say about Tanzanians? Students at African universities are frequently more educated than everyday citizens and universities tend to be located in urban areas, which can exacerbate opposition tendencies already present in these institutions. For professors and students, universities often represent a relative safe haven where they may express controversial political views. Indeed, this may be why one of the major limitations of this study—that incumbents sabotaging the opposition and domination of the police and military are apparently not very sensitive issues for University of Dar es Salaam students. Hence, one must be cautious not to claim too much for what these data say about Tanzanians. Still, I would expect potential biases in the survey sample identified above would make for a "hard test" and generally to work

---

[31]See Morton and Williams (2010: 237)

against the findings presented here. While students may generally support the opposition more, they are also more likely to be willing to pay the costs for being vocal about these views, thus reducing the difference between direct and less obtrusive techniques.[32] The same goes for any survey instrument effects that could arise by my use of a self-administered questionnaire because it conceals the identity of the survey subject from the researcher. The more commonly used approach for surveys in sub-Saharan Africa is face-to-face interviews, so we might expect the biases to be even greater in these surveys.[33]

6.2 Internal Validity: List Experiment[34]

I found in the list experiment that, for each of the three sensitive questions, the mean number of items respondents agreed with ranged between 2.15 and 2.35 for the control groups and over three for the treatment group. That the mean number of items of agreement in the control group hovered around these values is, in part, an artifact of design. Following Kuklinski et al. (1997b) and Glynn (2009), the treatment lists were designed to reduce the likelihood of "ceiling effects" by having two pairs of items that were somewhat conflicting. Ceiling effects were minimal and did not violate the assumptions underlying the list experiment.[35] This claim was validated by cross-checking the data with Glynn (2009)'s piecewise estimation procedure.[36] Still, a reading of the literature on list experiments shows that for 4 item control lists, the mean number items selected by respondents tends to range between 2.1 and 2.3. It could be that this is occurring because of some innate "circling bias" that individuals have, driving them to select a value independently of the items on a list. On one hand, such a tendency could arise due to burnout from engaging the more complex question format[37] or that respondents forget some of the items on the list and offer a value they think is a reasonable. We expect that recall effects are minimized simply because the questionnaire was self-administered. It could also be that there is a "midpoint response style" and that the populations which have been sampled for list experiments share some attribute that leads to a certain way of answering list items—in this case,

---

[32]It remains a possibility that conducting the experiment outside of a university could be an even "harder" test because response effects might surface in all three question formats. While my data do not permit me to make a claim on this issue, I am conducting a three round survey in Dar es Salaam before, during, after the 2010 national elections and expect that the data from these collection rounds will.

[33]While this argument has been made by many, Groves et al. (2004) approach this discussion in particular detail.

[34]A tip of the hat to Kevin Fridy, Macartan Humphreys, and Eric Kramon for comments I draw upon in this section

[35]Including pre-testing and using subgroup estimates from previous public opinion studies to gauge the expected rate of agreement with a particular statement.

[36]Available from author.

[37]On satisficing, see Krosnick (1991)

selecting a middle value.[38] Future work using list experiments will do well to vary the number of control items on lists—testing 4 control items with a fifth treatment, alongside surveys with 5 control items and a sixth treatment—to ensure these factors are not influencing results.

6.3 Internal Validity: Randomized Response Technique

I conducted pre-tests of the survey instrument with a number of students and found they indicated little trouble in completing the questionnaires on the basis of provided instructions within the questionnaire and a prompt delivered before distributing them. Still, I have limited confidence in results produced by the Randomized Response Technique.

What permits the RRT to mask an individual's response to a question is that they have randomly been assigned to answer a question for which there is a known value—that 37% of students who received questionnaires were female; half of the recipients were prompted to answer this innocuous question. Treatment assignment came by way of a picture that was stapled to the inside cover of each questionnaire which respondents were instructed to remove before returning the survey so that individual responses could not be identified. However, 134 of the 261 respondents (53.1%) submitted instruments with the pictures still attached. Either subjects were not concerned about the sensitivity of their responses—which based on other evidence, I find to be unlikely—or over half of them did not understand the instructions or the RRT itself. This may explain the rather surprising finding in Table 3 that a smaller proportion of respondents who received the randomized response technique believed that the ruling party sabotages its competition than those who received the direct question. While RRT did perform far better for the other two questions, I am reluctant to place much stock in these findings, given the problems surrounding the attached pictures on returned surveys.

Still, this presents a unique opportunity to assess the extent to which a non-probabilistic parameter to estimate baseline "yes" responses for the non-sensitive question—in this study, gender—operates in comparison to random events like coin-flipping, which may have a lot of noise. Of the 134 respondents returning surveys with pictures still attached, 82 of them were the picture associated with the non-sensitive item.[39] In

---

[38]Kieruj and Moors (2010) offer a review of the problems of "midpoint" and "extreme" response styles in surveys using numerical scales.

[39]This offers statistically significant evidence that subjects were systematically less likely to acknowledge assignment to the

other words, I can actually determine the proportion of respondents who were assigned the gender question AND responded that they were female. Table 5 presents this information for each sensitive question. Note how the returned pictures suggest that a fixed value like the percentage of respondents matches quite well with the estimate of 37% used in the previous sections. Moreover, we see that the greatest deviation from this fixed value was the second sensitive item, where the RRT technique unearthed unexpectedly lower estimates of the belief that incumbents interfere with opposition parties. It might be that there was some bias in the respondents who received this format and returned the questionnaire. Still, using a fixed value like the number of females in a classroom may provide an interesting way to avoid the statistical noise associated with probabilistic events like flipping a coin or a die which would pronounced with smaller sample sizes. While it may not be possible in all cases, other scholars may find this approach a more elegant way of administering randomized response technique.

INSERT TABLE 5 ABOUT HERE

7 Discussion

Poor survey data quality due to self-censoring has prevented social scientists from properly studying sensitive political topics. This paper provides evidence that list experiments and to a lesser extent randomized response provide a better way of measuring public opinion about such issues. It also identifies some potential costs associated with using them—particularly, generating group level data and employing more complex question formats. Still, given the remarkably poor data produced by direct questioning, it seems potential gain from using alternative techniques like list experiments and RRT far outweigh these costs. Moreover, as these formats become more widely used, scholars will develop more sophisticated means of analyzing such data.[40]

This paper joins a growing set of literature exploring techniques that reduce sensitivity bias issues in different political contexts. First of all, evidence from the study suggested that Randomized Response is an approach that might cause respondents too much confusion to be a reliable technique for studying

---

sensitive question ($\chi^2 = 8.44$, *df*=1)

[40]Corstange (2009), Glynn (2009) and Imai (2011) are pioneering works to this end.

political attitudes in sub-Saharan Africa. The remaining pictures left on many of the instruments and the counter-intuitive findings with respect to the question about opposition sabotage suggested that the technique suggests that the technique might be too complex for university students to understand and implement. More research must be done before writing off the method's applicability for political attitudes in sub-Saharan Africa, but it is quite reasonable to expect these comprehension problems to be even greater with lesser-educated or illiterate subjects often found in the sub-continent. A recent telephone survey using randomized response conducted in the United States showed major implementation problems due to respondent error, giving us even further grounding to this concern.[41]

The findings with respect to the list experiment were more promising, though more research will also be done in these contexts in order to ensure they perform well with a non-university sample. Nonetheless, improved design and analysis of list experiments and their application to a growing variety of political phenomena and comparative settings stand to expand our ability measure and understand sensitive political topics. List experiment design insight—including reducing the likelihood of "ceiling" and "baseline" effects that limit to ability of the method to mask individual responses, increasing precision with smaller sample sizes through the double list experiment, and debate about whether or not control lists should be "open" and ask respondents to identify *which* control items they support—all represent fruitful avenues. Similarly, tools are now available that shift the analytical power of list experiments from univariate, group comparisons to group and individual level multivariate analyses.[42] Finally, list experiments have been applied to an emerging number of regions throughout the developing world in order to understand an array of political phenomena unique to under-developed and semi-democratic contexts. These studies include analyses of vote-buying in Latin America,[43] and voting rights extension in the Middle East[44] New applications of list experiments offer emerging opportunities to understand better how they work and the sorts of possibilities and potential limitations they have for better understanding sensitive political attitudes.

---

[41]Holbrook and Krosnick (2010a)

[42]See Corstange (2009), Glynn (2009), Imai (2011)

[43]Gonzales-Ocantos et al. (2011)

[44]Corstange (2009)

References

Abramson, P. R. and Claggett, W. (1984). Race-related differences in self-reported and validated turnout. *Journal of Politics*, 46:719–38.

Achen, C. H. (1975). Mass political attitudes and the survey response. *American Political Science Review*, 69(4):1218–31.

Acock, H. (2005). Working with missing values. *Journal of Marriage and the Family*, 67:1012–28.

Althaus, S. L. (1998). Information effects in collective preferences. *American Political Science Review*, 92(3):545–58.

Barkan, J. D. (1994). *Beyond Capitalism vs. Socialism in Kenya and Tanzania*. Lynne Rienner Publishers, Boulder, CO.

Bernstein, R., Chadha, A., and Montjoy, R. (2000). Overreporting voting-why it happens and why it matters. *Public Opinion Quarterly*, 65:22–44.

Bogaards, M. (2004). Counting parties and identifying dominant party systems in africa. *European Journal of Political Research*, 43:173–97.

Cassel, C. A. (2003). Overreporting and electoral participation research. *American Politics Research*, 31:81–92.

Chaligha, A., Mattes, R., Bratton, M., and Davids, Y. D. (2002). "uncritical citizens" or "patient trustees": Tanzanian's views of political and economic reform. *Afrobarometer Working Paper Series*, 18.

Chaudhuri, A. and Mukerjee, R. (1988). *Randomized Response: Theory and Techiques*. CRC Press, Boca Raton, FL.

Converse, P. E. (1964). Constraint among elements. In Apter, D. E., editor, *Ideology and Discontent*. The Free Press, New York.

Corstange, D. (2009). Sensitive questions, truthful answers? modeling the list experiment multivariately with listit. *Political Analysis*, 17:45–63.

Coutts, E. and Jann, B. (2008). Sensitive questions in online surveys: Experimental results for randomized response technique (rrt) and the unmatched count technique (uct). Unpublished Manuscript.

Duff, B., Hanmer, M. J., Park, W.-H., and White, I. K. (2007). Good excuses: Understanding who votes with an improved turnout question. *Public Opinion Quarterly*, 71(1):67–90.

Fox, J. A. and Tracy, P. E. (1986). *Randomized Response: A Method for Sensitive Surveys*. Sage Publications, Thousand Oaks, CA.

Gingerich, D. W. (2010). Understanding off-the-books politics: Conducting inference on the determinats of sensitive behavior with randomized response surveys. *Political Analysis*, 18:349–380.

Glynn, A. N. (2009). On the design of statistical truth serum: Getting more from the list experiment with sample size calculation, design principles, and the probablisitc imputation of socially undesirable responses. Unpublished Paper.

Gonzales-Ocantos, E., Kiewet de Jonge, C., Melendez, C., Osorio, J., and Nickerson, D. (2011). Vote buying and social desirability bias: Evidence from nicaragua. *American Journal of Political Science*, Forthcoming. Unpublished Manuscript.

Groves, R. M., Floyd J. Fowler, J., Couper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R. (2004). *Survey Methodology*. John Wiley and Sons, Hoboken, NH.

Holbrook, A. L. and Krosnick, J. A. (2010a). Measuring voter turnout by using the randomized response technique: Evidence calling into question the method's validity. *Public Opinion Quarterly*, 74(2):328–343.

Holbrook, A. L. and Krosnick, J. A. (2010b). Social desirability bias in voter turnout reports. *Public Opinion Quarterly*, 74(1):37–67.

Hyden, G. (1980). *Beyond Ujamaa in Tanzania: Underdevelopment and an Uncaptured Peasantry*. University of California Press, Berkeley, CA.

Hyden, G. (1999). Top down democratization in tanzania. *Journal of Democracy*, 10(4):142–55.

Imai, K. (2011). Multivariate regression analysis for the item count technqiue. *Journal of the American Statistical Association*, Forthcoming.

Kane, J. G., Craig, S. C., and Wald, K. D. (2004). Religion and presidential politics in florida: A list experiment. *Social Science Quarterly*, 85:281–93.

Kieruj, N. D. and Moors, G. (2010). Variations in response style behavior by response scale format in attitude research. *International Journal of Public Opinion Research*, 22(3):320–42.

King, G., Honaker, J., Joseph, A., and Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review*, 95(1):49–69.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5:213–36.

Kuklinski, J., Cobb, M. D., and Gilens, M. (1997a). Racial attitudes and the 'new south'. *Journal of Politics*, 59:323–49.

Kuklinski, J., Sniderman, P. M., Knight, K., Piazza, T., Tetlock, P. E., Lawrence, G. R., and Mellers, B. (1997b). Racial prejudice and attitudes toward affirmative action. *American Journal of Political Science*, 41(2):402–19.

Levitsky, S. and Way, L. A. (2010). *Competitive Authoritarianism: Hybrid Regimes After the Cold War*. Cambridge University Press, Cambridge.

Martinez, M. D. and Craig, S. C. (2010). Race and 2008 presidential politics in florida: A list-experiment. *The Forum*, 8(2):1–14. Forthcoming.

Mkapa, B. W. (2003). *Uwazi na Ukweli: Rais wa Watu Anazungumza na Wananchi: Kitabu cha Kwanza*. Mkuki na Nyota Publishers, Dar es Salaam, Tanzania.

Mmuya, M. and Chaligha, A. (1994). *Political Parties and Democracy in Tanzania*. Dar es Salaam University Press, Dar es Salaam, Tanzania.

Morton, R. B. and Williams, K. C. (2010). *Experimentation in Political Science and the Study of Causality: From Nature to the Lab*. Cambridge University Press, Cambridge. New York University. Book Manuscript.

Msekwa, P. (2006). *Reflections on the First Decade of Multiparty Politics in Tanzania*. Hanns Seidel Foundation, Dar es Salaam, Tanzania.

Rayburn, N. R., Earleywine, M., and Davison, G. C. (2003). An investigation of base rates of anti-gay hate crimes using the unmatched-count technique. *Journal of Aggression, Maltreatment and Trauma*, 62(2):137–53. Unpublished Manuscript.

Schedler, A. (2006). The logic of electoral authoritarianism. In Schedler, A., editor, *Electoral Authoritarianism: The Dynamics of Unfree Competition*, Boulder, CO. Lynne Rienner Publishers.

Sniderman, P. M. and Carmines, E. G. (1997). *Reaching Beyond Race*. Harvard University Press, Cambridge, MA.

Soloman, J. S. K. J. K. W. M. G. (2007). Estimating illegal resource use with randomized response technique. *Human Dimensions of Wildlife*, 12:75–88.

Streb, M. J., Burrell, B., Fredrick, B., and Genovese, M. A. (2008). Social desirability effects and support for a female presidential candidate. *Public Opinion Quarterly*, 72:76–89.

TEMCO (2006). *The 2005 General Elections in Tanzaia; Report of the Tanzania Election Monitoring Committee*. TEMCO, Dar es Salaam, Tanzania.

Tsuchiya, T., Hirai, Y., and Ono, S. (2007). A study of the properties of the item count technique. *Public Opinion Quarterly*, 71(2):253–72.

Warner, S. L. (1965). Randomize response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–9.

Wimbush, J. C. and Dalton, D. R. (1997). Best rate for employee theft: Convergence of multiple methods. *Journal of Applied Psychology*, 82(5):756–63.

Tables and Figures

Table 1: Matrix of Split Thirds Design

	Group 1	Group 2	Group 3
Opposition Violence	Direct	List Exp.	RRT
Number of Observations	*89*	*86*	*79*
Incumbent Sabotage of Opposition	List Exp.	RRT	Direct
Number of Observations	*90*	*87*	*80*
Incumbent Control of Police/Military	RRT	Direct	List Exp.
Number of Observations	*87*	*83*	*82*
List Exp. Control, Opposition Violence ($N = 161$)	Yes	No	Yes
List Exp. Control, Incumbent Sabotage ($N = 166$)	No	Yes	Yes
List Exp. Control, Police/Military ($N = 170$)	Yes	Yes	No

Table 2: Oppositions Sometimes Must Use Violence

	Control	Treatment[†]	RRT[‡]	Direct Question
Mean Number of Agreed List Items	2.15	3.05***		
Std. Error	(0.07)	(0.10)		
Total Proportion of "Yes" for RRT			59.49%	
Belief Incumbents Sabotage Opposition		86.26%	81.98%	41.57%
95% Confidence Interval*		[61.82, 100.00]	[60.96, 100.00]	
N	161	86	79	89

† * $p < .05$ ** $p < .01$ *** $p < .001$, Welch's t-test
‡ Control Group 37% "Yes"
* Confidence Interval Truncated at 100%

Table 3: Incumbents Sabotage Opposition

	Control	Treatment[†]	RRT[‡]	Direct Question
Mean Number of Agreed List Items	2.33	3.32***		
Std. Error	(0.08)	(0.12)		
Total Proportion of "Yes" for RRT			48.28%	
Belief Incumbents Sabotage Opposition		90.20%	59.56%	82.93%
95% Confidence Interval*		[61.56, 100.00]	[38.58, 100.00]	
N	166	90	87	82

† * $p < .05$ ** $p < .01$ *** $p < .001$, Welch's t-test
‡ Control Group 37% "Yes"
* Confidence Interval Truncated at 100%

Table 4: Ruling Party Controls Policy/Military

	Control	Treatment[†]	RRT[‡]	Direct Question
Mean Number of Agreed List Items	2.25	3.05***		
Std. Error	(0.08)	(0.12)		
Total Proportion of "Yes" for RRT			58.62%	
Belief Incumbent Controls Police/Military		79.58%	80.24%	79.52%
95% Confidence Interval*		[52.22, 100.00]	[59.52, 100.00]	
N	170	82	87	83

† * $p < .05$ ** $p < .01$ ***$p < .001$, Welch's t-test
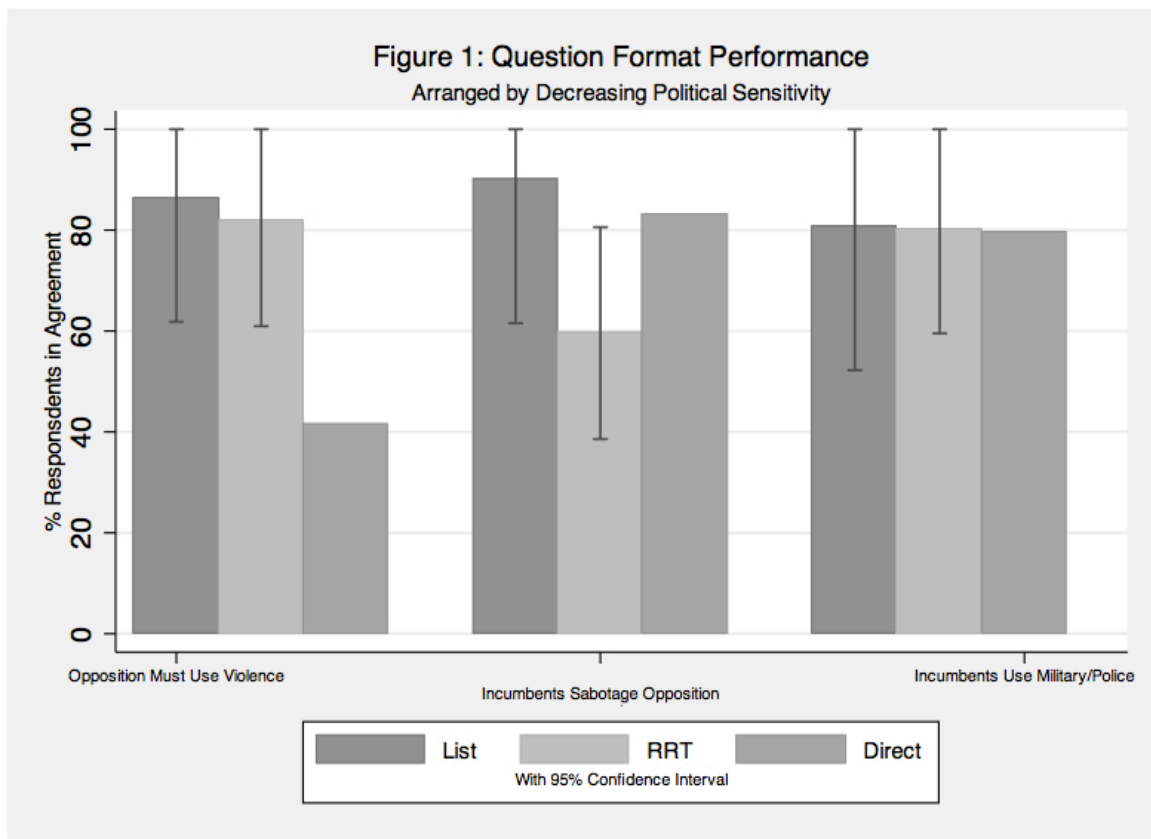‡ Control Group 37% "Yes"
* Confidence Interval Truncated at 100%

Figure 1: Question Format Performance
Arranged by Decreasing Political Sensitivity

Table 5: Instruments with Returned Picture

	Opposition Violence	Incumbent Sabotage	Control Police/Military
Total Instruments Returned with Picture	41	43	50
Instruments with non-Sensitive Picture	26	30	26
% "Yes" for Female Question	38.46%	30.00%	34.62%
% "Yes" for Sensitive Question	60.00%	69.23%	79.17%

Appendix

Treatment Lists for List Experiments

"Below, I provide a list of different political attitudes that Tanzanians sometimes agree with:

· In Tanzania, nation is the strongest form of identity.
· The cost of basic goods like grain is too high.
· Tanzania has politically performed better than its neighbors.
· Daladala drivers need greater training in order to reduce risks they post on the road.
· The reason opposition parties are unsuccessful in elections is because the ruling party sabotages their attempts to compete.

How many of the items do you agree with? Do not indicate which items, just how many of the five items you agree with. Please circle one of the numbers below.

"Below, I provide a list of different political attitudes that Tanzanians sometimes agree with:

· It's time to leave the legacy of *Ujamaa* behind.
· Private media should be restricted by the government unless they are properly trained.
· My community looks up to me as a representative of them.
· Students should receive more support from the government to help pay for university.
· The military and police primarily carry out the agenda of the ruling party.

How many of the items do you agree with? Do not indicate which items, just how many of the five items you agree with. Please circle one of the numbers below.

Computing Estimates for Randomized Response Technique

Drawing from Fox and Tracy (1986) and Chaudhuri and Mukerjee (1988),[45] the overall probability of a "yes" response λ can be expressed as the probability of a "yes" response given assignment to the sensitive question π_A, the probability of a "yes" given assignment to the non-sensitive item π_Y, and probability being assigned to the sensitive item $P \in [0, 1]$

$$\lambda = P\pi_A + (1 - P)\pi_Y \tag{1}$$

Thus, where $\hat{\lambda}$ represents the total proportion of "yes" responses, we have an unbiased estimator of the proportion of "yes" responses for the sensitive question π_A

$$\hat{\pi_A} = [\hat{\lambda} - (1 - p)\hat{\pi_y}]/p \tag{2}$$

---

[45]The equations presented above appear on pages 18-21 and 16-18 of their works respectively.

with variance

$$Var(\hat{\pi}_A) = \hat{\lambda}(1 - \hat{\lambda})/np^2 \tag{3}$$

In this case, the probability of assignment to the sensitive item is $p = .5$. Moreover, rather than following many applications of RRT that use randomizing events to infer the parameter π_Y (like a second coin flip or die roll) , I instead use a parameter for which the true value is known—the percentage respondents who are female. Thus, for the proportion of "yes" responses to the nonsensitive item is $\hat{\pi}_Y = .37$.