

Competence or Charisma?  
The effects of measurement on why some leaders' traits matter more than others

Amanda Bittner  
University of British Columbia  
[bittnera@interchange.ubc.ca](mailto:bittnera@interchange.ubc.ca)

Prepared for Presentation at the Canadian Political Science Annual Meeting, University of Saskatchewan, Saskatoon, SK, May 30-June 1, 2007.

This paper was made possible as a result of the support of the Social Sciences and Humanities Research Council of Canada, in the form of a doctoral fellowship. I would like to thank Fred Cutler, Richard Johnston, and Scott Matthews, for their comments and advice during the preparation of the paper. Of course, none of these individuals are responsible for any arguments or errors in this paper.

## Abstract

Comparative study of the impact of party leaders and their personality traits on election outcomes is diverse and non-cumulative—it has followed a number of directions, and the literature is conflicted in almost all of these areas. One of the main branches within the leaders literature focuses on the evaluation of personality traits, and scholars are unable to agree about which traits voters take note of; whether traits actually affect election outcomes; if they do which are more important; and why or how traits are evaluated in the first place. When one looks more closely at the literature, it becomes clear that what is needed is a larger comparative analysis that incorporates election studies across time and space, in order to be able to make more definitive conclusions about the role of leader evaluations. However, one of the challenges to this endeavor stems from the nature of the data available. Respondents are asked to evaluate different personality traits in different election studies over time, and often even where trait continuity does exist, the question format changes, making comparability over time and across space very difficult. This paper examines the impact of question format, using pooled data from election studies conducted in two countries, Canada and the United States. The analysis suggests that question format does indeed have an effect on respondents' evaluations of party leaders, but more research is needed before we can make conclusions about the implications of these effects for the comparative study of the role of leaders.

## Introduction

Media coverage of recent elections around the world points to the important role played by party leaders in the election process. Leaders not only provide a “face” for the party, but election wins and losses are also often credited to or blamed on leaders as individuals. Examples abound. Putting aside issues surrounding the counting of ballots and the electoral college for a moment, the 2000 US election is often seen as a face-off between Bush and Gore, with Gore losing as a result of perceptions of his honesty and his wooden demeanor. In the five-party competition seen in the 2000 Canadian election, Stockwell Day's lackluster performance at the polls is often credited to suspicions about his personality, his motives, and his slick appearance. Academics have charted the particular “horserace” style of campaign coverage, and have noted substantial effects on election campaigns, from the priming of party leadership (Mendelsohn 1994; 1996), to the impact coverage has on campaign contributions and finances (Mutz 1995), to the media's focus on certain types of polls and information over others (Lewis and Wahl-Jorgensen 2004). The media focus on leaders leads one to assume that perceptions of party leaders play an important role in determining vote choice, full stop.

While the media may have decided that leaders matter, the academic literature on the role of party leaders in election outcomes is conflicted and inconclusive. One of the main branches within the leaders literature focuses on voter evaluation of personality traits (e.g. honesty, integrity, intelligence, and trustworthiness, to name a few). Scholars are unable to agree, however, about which traits voters take note of; whether traits actually affect election outcomes; if they do, which are more important; and why or how traits are evaluated in the first place. While this lack of

agreement is wonderful for scholars looking for something to do, it also leaves our understanding of the role of leader evaluations less than complete.

When one looks more closely at the literature, however, it becomes clear that part of the lack of agreement probably stems from the nature of the analyses conducted to date. Generally, analyses of the role of leaders tend to be based upon a single-country cross-sectional analysis of a single election (Brown et al. 1988; Conover and Feldman 1989; Crewe and King 1994; Mendelsohn 1994; Stewart and Clarke 1992). Longitudinal analyses have been conducted, but these results are also largely focused on single-country studies (Clarke, Ho, and Stewart 2000; Gant and Sigelman 1985; Gidengil et al. 2000; Graetz and McAllister 1987a; Hudson 1984; Stokes, Campbell and Miller 1958). There has been some comparative and longitudinal work, but these studies make up a substantially smaller proportion of the field (Banducci and Karp 2000; Bean 1993; Bean and Mughan 1989; Graetz and McAllister 1987b). How conclusive can we really expect things to be with a series of disconnected studies with different findings? While one might be tempted to find fault with these scholars, the bigger problem really is the lack of continuity over space and time in the election studies that all of this research is based on. Simply put, different traits are included in different election studies over time (even in the same country), and often, even where trait continuity does exist, the question format changes, making comparability over time and across space extremely difficult. Our ability to draw substantive conclusions is, consequently, limited.

The observation that question formats may change even when there is trait continuity really begs the question of how much of a difference question format actually makes in affecting the evaluation of leader traits. If question format has a negligible impact on the evaluation of traits and the impact of these traits on overall leader evaluations, then the number of comparative research opportunities increases, and there is potential for scholars to cast their nets wider in order to develop a more encompassing theory of the role of leader evaluations. If, however, question format does have an impact on evaluations, then we must think more closely about how we might embark on a larger comparative project.

This study looks at the effect of question format in the context of two countries, Canada and the United States, and four election studies overall: the 1993, 1997, and 2000 Canadian Election Studies, and the 2000 National Annenberg Election Study in the US. The paper proceeds as follows: first it moves beyond these two countries to assess the diversity of traits found in election studies across time and space, as well as the diversity of question formats over time. Second, it moves on to an analysis of the impact of question format in both the Canadian and Annenberg Election Studies. Finally, it will lay out the limitations and implications of the study. This paper finds that while question format does have an impact on respondents' evaluations of leaders' traits, the effects are mixed and inconclusive. More research is needed in order to better understand the impact of question format.

### **The Diversity of Traits and Questions Across Time and Space**

This project began with a survey of election studies to first determine the number of studies that include questions about leader traits, and then assessed the level of continuity in specific traits questions across space and time. As such, the survey questionnaires of 93 election studies (all of

which included at least one question about leaders) were examined from 16 countries.<sup>1</sup> The list of election studies incorporated is not exhaustive, but includes only those election studies for which questionnaires could be found in English or French, the author's language skills being limited to those two. Thus it is important to note that the following estimates of frequency of question types and trait types are slightly conservative, as they do not represent a complete list of all election studies that have been conducted. Of these 93 election studies, 59 incorporate questions involving evaluations of leader traits, and 46 of these 59 studies probe for evaluations of traits in a closed-ended format. Table 1 (see Appendix A) lists all of the traits that have been included in these election studies over time, and also lists the election studies and years in which the particular trait has been asked. As the Table illustrates, there are over 154 separate specific "trait wordings" which have been asked in all of the 46 election studies incorporating closed-ended trait questions. While some of these can be grouped together, leaving us with a smaller total number of traits, the number of separate traits is still well over 100. This does suggest that it might be difficult to generalize about voters' evaluations of leaders' personality traits either longitudinally or cross-nationally, since there is substantial variety in the types of traits that respondents are asked to evaluate.

While there is substantial variation in the specific traits that are included in election studies there is also considerable overlap, as Table 1 also indicates. Most commonly asked is "strength of leadership," included in 27 election studies, followed by "responsiveness to ordinary values," asked in a number of different formats, 22 times in total. Other popular trait types are "honest" (18 times), "knowledgeable" (17 times), "intelligent" (16 times), "compassionate" (15 times), "moral" (14 times), "trustworthy" (13 times), and "arrogant" (11 times). These commonalities make room for comparison over time and space, but unfortunately the way in which respondents are prompted to evaluate these traits varies substantially as well. For example, some studies ask respondents to identify the leader who best fits a particular trait (as in Britain in 1983, Canada in 2000, and Israel in most years), while others ask respondents to indicate the extent to which a trait describes a leader on a scale (1 to 7, 0 to 10, -5 to +5, etc.) in which (usually) the higher number, the better the trait is said to describe a leader. Others still provide respondents with sets of trait "opposites" (e.g. strong/weak; can work on a team/cannot work on a team) on a scale, where each end of the scale represents one of the traits. Respondents are then asked to indicate which number on the scale best represents the leader's fit in the spectrum. The closer the number to one of the traits, the better that trait fits the leader. This format has tended to be particularly popular in Israel, Germany, and Britain.

The format of trait questions is a factor that must be taken into consideration when thinking about how one might conduct a comparative analysis across space and time. For example, comparing evaluations of leaders' traits in Canada in 1997 versus 2000 presents a challenge, since the question formats in the two years were so different. In 1997, Canadian respondents were asked trait questions in the following format:

---

<sup>1</sup> These surveys include Australia (1987-2004); Britain (1974-2005); Canada (1965-2006); Finland (1991); France (1967, 1968, 1988, 1997); Germany (1969-1998); Hong Kong (1998-2000); Israel (1969-2003); Italy (1968-1972); Japan (1967); The Netherlands (1967-1998); New Zealand (1993-2002); Sweden (1956-1991); Czech Republic (1996); Russia (1995-96); US-NES (1952-2004); and US-Annenberg (2000-2004).

*(Q11a-e) Now we'd like to get your impressions of the party leaders. I would like you to tell me how well the following words fit each leader. What about Jean Chretien. Does 'arrogant' describe Jean Chretien very well, fairly well, not very well, or not at all?*

Meanwhile, in 2000, Canadians were asked a very different question format:

*(Qe1-5) "Which party leader would you describe as arrogant?"*

As is fairly obvious, these two questions are not exactly the same, and trying to compare trait evaluations in these two years poses some methodological challenges. Before embarking on a large-scale comparative study, it is important to establish what effects, if any, these question formats may have on evaluations of party leaders.

Other types of survey question format effects are also worth understanding. For example, during the 2000 American Election, the National Annenberg Election Study randomized the way in which respondents were asked to evaluate the traits of Presidential candidates: specifically, the way that traits were clustered varied for each group. One half of the respondents was asked to evaluate all of the traits for one leader before moving on to the same list of traits for the second leader (labelled as "traits within candidates"). The other half of the sample was asked to evaluate a single trait for both leaders before moving on to the next trait (labelled for this study as "candidates within traits"). While the majority of election studies (64% of the election studies surveyed in this study) tend to follow the format of "traits within candidates," there is some variation, and the Annenberg study allows us to evaluate the effects of trait grouping on leader evaluations. Both trait grouping and question format vary over time and space, and in order to conduct a broad comparative analysis of the role of the evaluation of leader traits, it is important to understand the effects of survey and question format. This study is an attempt to begin to unpack some of these effects.

## **The Canadian Election Studies**

As indicated earlier, even where trait continuity exists in the Canadian Election Studies, question format varies considerably. In the eight Canadian Election Studies which include closed-ended questions about leaders traits, there are four different question formats. Table 2 lays out the traits, question formats, and party leaders in these eight election studies. In 1968, the question format was as follows:

*Q59a - Now I am going to read you a list of things often said about Mr. Trudeau. I would like your reaction to them. In each case please tell me whether you agree with the statement or whether you disagree. Mr. Trudeau is highly intelligent...etc.*

Thus there were two categories of responses in this study, agree or disagree.

In contrast, there were many categories of response in the 1984 election study. Respondents were asked to evaluate leaders' traits on a scale of 1-7, with the question format as follows:

*QF8 - Now we'd like to know your impressions of what certain political leaders are like. I'll read a word or phrase, and I'd like you to tell me how well it fits or describes each of*

*the three leaders shown on this card. The more you think the word or phrase describes a leader, the closer your answer should be to 7. The less it fits, the closer your answer should be to 1. If you have no idea at all about how a word or phrase fits a leader, tell me and we'll go on to the next one. First, in your opinion, how well does the word "arrogant" describe Turner? Which number on this scale gives the best idea of how you see him?*

The period from 1988 to 1997 represents the greatest level of consistency in both traits and question format. From 1988 to 1997, respondents were asked about the extent to which they felt that a trait described a party leader.

*Qd3 - Now we'd like to know about your impressions of the party leaders. I am going to read a list of words and phrases people use to describe political figures. After each one, I would like you to tell me how much the word or phrase fits your impressions. How much would you say 'intelligent' fits your impression of Brian Mulroney: a great deal, somewhat, a little, or not at all?*

This format, while not quite as precise as the scale format, still provides respondents with the ability to make finer gradations in their evaluations of leader traits. In contrast, in 2000, question format changed substantially and respondents were more constricted in their options of how to evaluate leaders. In this election, individuals were asked to choose a leader who best represented a given trait:

*Qer-5 - Which party leader would you describe as arrogant? Trustworthy? Having new ideas?*

While it was not listed in the question, responses of "all" or "none" were also accepted, but this question generally provides fewer response options.

In 2004 and 2006, the question format changed again, this time returning back to the scale format similar to that used in 1984, only this time with more points on the scale. The question format is as follows:

*Qh1a-d-h2a-d - We want to ask you how honest each party leader is. On a scale from 0 to 10 where 10 means very honest and 0 means very dishonest, how honest do you think <name of leader> is?*

As the above description and Table 2 illustrate, making observations about the impact of leaders' traits in Canadian elections requires some consideration of the methodological challenges involved in comparing trait evaluations from one year to the next, given the substantial changes in question format over time (see Johnston 2002 and Nakai 2003 for additional discussion of methodological issues which arise). As such, it is important to attempt to determine what effect these question format changes may have on respondents' evaluations of leaders, and to incorporate these considerations into future analyses.

Part of the problem is that even attempting to unpack these question format effects is difficult—leaders change, traits change, and different election years involve different campaign dynamics even if the same leaders are present—making case selection and election study comparability an issue. In 1993, 1997, and 2000, respondents were asked to evaluate three of the same traits, and

there was also some continuity in leadership (Jean Chretien ran for the Liberal Party in 1993, 1997, and 2000, and Alexa McDonough and Gilles Duceppe ran for the NDP and Bloc Quebecois, respectively, in the latter two years). This level of similarity allows us to isolate the effects of question format, since 1993/1997 and 2000 make use of two different types of question format.

As described above, in 1993/1997, individuals were asked how well a given trait described a leader. Respondents could pick the category that they felt best represented the fit between the trait and the leader. In 2000, however, the format was substantially different, and respondents were asked to *choose* the leader that they felt fit a particular trait. What effects might we expect these question formats to have?

It is hypothesized that the act of choosing is important: if a respondent actually chooses a leader from among the others as the best representation of a given trait, then that choice is likely to have a stronger effect on overall evaluations of that leader than simply fitting the leader within some broad descriptive categories, as in 1993/97. In order to choose a leader, the respondent is more likely to feel strongly about that trait description – a kind of “top of the head” effect – thus the link between that trait and the overall evaluation of the leader should be stronger. Thus for example, it is expected that those who choose Chretien to fit the description of “arrogant” will have lower overall ratings of him than those individuals who agree that he is arrogant. Similarly, those who choose him to fit the descriptions of “trustworthy” and “compassionate” will have more positive overall feelings towards him than those individuals who simply agree that he possesses those traits.

In order to test this hypothesis, data from the election studies from 1993, 1997, and 2000 were pooled and analyzed. All three election studies include a thermometer rating for each of the party leaders, a measure that is used quite regularly to represent overall evaluations of party leaders, and which will act as the dependent variable in this analysis. The question for all three years is nearly identical, and reads as follows (this version from the 1997 CES):

*Qdia-e – Now we’re going to ask you how you feel about the party leaders using a scale from 0 to 100. 0 means you really dislike the leader and 100 means you really like the leader. You can use any number from 0 to 100. How do you feel about ---?*

As Table 2 illustrates, 1997 and 2000 have three party leaders in common (Jean Chretien, Alexa McDonough, and Gilles Duceppe) as well as three traits in common (arrogant, trustworthy, and compassionate). Meanwhile, the question format changes in 2000, providing an opportunity to look at its effects on overall evaluations for the same traits for the same leaders in consecutive elections. There are some complications, however. The 1997 and 2000 elections were very different elections, thus any changes in evaluations that we see from one year to the next which might be attributed to the question format might also be attributed to circumstances particular to the election. It is for this reason that the 1993 election study was included, to extend the analysis further for Chretien, in order to make an effort to tease out question effects from overall year/election effects. This will be discussed further below.

In this analysis, thermometer evaluations for the above three leaders were regressed on the three traits, in addition to a variable representing question format, interactions between the traits and the format variable, and a battery of demographics standard in Canadian analyses of voting and

public opinion, as well as party identification.<sup>2</sup> Table 3 illustrates the results of these regression analyses, for evaluations of Alexa McDonough and Gilles Duceppe, and Table 4 reports results for overall evaluations of Jean Chretien.

As the tables indicate, while the effects of traits on overall ratings were significantly different for each of the two question formats, the results do not conform completely with the hypothesized relationship (that the “choice” would involve a stronger relationship between the trait and the overall rating). In fact, the hypothesis is partially confirmed, and in the same way for all leaders: Tables 3 and 4 suggest that the impact of the “choice” on thermometer ratings is strongest for the negative trait (arrogant) while the opposite is true for the positive traits—that is, the impact of the positive traits is stronger when the question was posed in the other (extent of agreement) format.

When we look at the first column of Table 3, for example, we can see this pattern. Those who felt that Alexa McDonough was arrogant – those who chose her from among all of the leaders – gave her overall thermometer ratings nearly 13 points lower on a scale of 0-100 than those who did not feel she was arrogant. For those presented with the other question format, the impact of feeling that arrogant described McDonough “a great deal,” was slightly smaller – respondents gave her overall thermometer ratings just over 10 points lower (coefficient is -0.102, standard error is 0.027) than those who felt that arrogant did not describe McDonough at all.<sup>3</sup> The relationship between the question format and the impact of the trait and the overall thermometer rating is reversed for the two positive traits. That is, a positive trait evaluation has a larger impact on overall thermometer ratings in the “extent of agreement” format than in the choice format. Still looking at evaluations of Alexa McDonough, those individuals who felt that she was trustworthy in the “choice” format gave her ratings just over 14 points higher on the 0-100 thermometer scale than those who did not feel she was trustworthy, while those who felt that trustworthy described McDonough “a great deal” gave her thermometer ratings nearly 34 points higher (coefficient is 0.337, standard error is 0.036) on the 0-100 scale than those who did not feel that trustworthy described her at all. This is a substantial difference. The same pattern exists for evaluations of the trait compassionate – those who chose McDonough as being compassionate gave her ratings just over 7 points higher on the thermometer scale, compared to those who did not think she was compassionate, while those who felt that compassionate described McDonough “a great deal” gave her ratings over 10 points higher than those who did not feel compassionate described her at all.

The second column in Table 3 illustrates the results of the analysis for Gilles Duceppe, which further illustrate this difference in relationship between question format and the impact of trait evaluations for positive and negative traits. Those who chose Duceppe as being arrogant gave him overall thermometer ratings that were nearly 14 points lower than those who did not feel he was arrogant, while those who felt that arrogant described Duceppe “a great deal” gave him thermometer ratings 5 points lower (coefficient is -0.053, standard error is 0.031) than those who

---

<sup>2</sup> See Appendix B for complete variable and coding details.

<sup>3</sup> This coefficient has been determined by obtaining the linear combination of the coefficients of both the trait (in the choice format) and the interactive term, using the STATA command “lincom.” This method is preferable to a rough calculation of the effect of the interaction, because it has the added benefit of combining the standard errors as well, thus allowing us to know the extent of the statistical significance of the combined coefficients. Further, even if the interaction coefficient does not achieve traditional levels of statistical significance, the linear combination of the two coefficients may achieve statistical significance.

didn't think arrogant described him at all. In contrast, for the positive traits, the "extent of agreement format" showed a stronger relationship between trait evaluations and thermometer ratings. Those who chose Duceppe as being trustworthy gave him thermometer ratings 15 points higher than those who did not choose him, while those who felt that trustworthy described Duceppe "a great deal" gave him thermometer ratings nearly 41 points higher (coefficient is 0.409, standard error is 0.040) than those who did not think that trustworthy described Duceppe at all. Again, this difference is very large. The coefficient for compassionate in the extent of agreement format was the same size as its standard error.

In discussing these results, no reference has been made to the question format dummy variable. Individuals who were presented with the 1997 "categories of agreement" format were coded 1 and those who were presented with the "choice" format were coded 0. It does not make much sense to talk about this variable on its own because of the difficulties in disentangling the question format effect from the "year" or "election" effect. The format variable by itself contains all of the overall year effects, leaving the interactions to tell us about the differential impact of each trait, by question format. It is for this reason that the focus has been on the traits on their own and the interactions with the question format variable. However, in order to provide some sort of control for "year" effects, the analysis was extended back by one election year for Jean Chretien, who had also been the leader in the 1993 election, in which the same traits and same format existed in the CES as it did in 1997. Table 4 illustrates the results of this analysis, in which a dummy was included for the year 1997 to be sure that we have a separate intercept for each of the elections, and to be able to further isolate the effects of the question format.<sup>4</sup> The only other change for this year was that Party Identification for the Bloc Quebecois was removed, as individuals were not prompted to claim an identification with that Party until the 1997 CES.

As Table 4 illustrates, again the data do not fully support the hypothesis that "choosing" a leader for each trait leads to a stronger effect on thermometer ratings. Similarly to evaluations of the other two leaders, it is only perceptions of Chretien's arrogance that lead to this pattern. Those who "chose" Chretien as being arrogant gave him ratings 17 points lower on the thermometer scale than those individuals who did not believe he was arrogant. The strength of this relationship decreased for those who felt that arrogant described him "a great deal," and they gave him ratings approximately nearly 8 points lower (coefficient is -0.077, standard error is 0.011) than those who did not think he was at all arrogant. The opposite occurred for the other two traits, where the impact of feeling that trustworthy (coefficient is 0.301, standard error is 0.014) and compassionate (coefficient is 0.144, standard error is 0.014) described Chretien "a great deal" on thermometer ratings was larger than it was among those who "chose" Chretien for those two traits. It is also important to note that this relationship holds even when we include a dummy variable for "year," thus giving us additional faith in the models for evaluations of McDonough and Duceppe.

These results suggest that question format has an impact on evaluations. However, the results are inconclusive, and the hypothesis that "choosing" a leader who fits a trait will have a stronger impact on overall thermometer ratings was only partially supported, as the data suggest that this relationship holds true for the negative trait "arrogant" but the opposite occurs for the two positive traits. That the pattern held for all leaders and all traits is particularly interesting, in that it suggests that perhaps this is not simply an anomaly. Is there something about negative traits that

---

<sup>4</sup> The inclusion of a year dummy for 1997, however, did not make a substantial difference on any of the coefficients, and the coefficient for the year variable itself did not reach traditional levels of statistical significance.

makes the impact of choosing greater while positive traits have a larger impact when we are asked to evaluate the extent to which they fit a given leader? More research is needed.

Regardless of whether the predicted relationship was supported in all circumstances, the results do clearly suggest that the question format affects overall evaluations – in some circumstances trait evaluations in the “choice” format more strongly affect thermometer ratings, and in others, the opposite holds true. These results suggest that survey design should consider the “type” of question being asked in comparison with the type of information sought, and researchers using these data ought to take into consideration the different question types when conducting their analyses. More research is needed, however, in order to provide a more conclusive account of the effects of question type. Extending the project to look at other national election studies may help, particularly if there is greater leader or trait continuity in these election studies, allowing us to better isolate the effects of question format.

### National Annenberg Election Study

The results from the Canadian analysis provide initial indications that the question format can affect the way in which perceptions of leader traits influence overall leader evaluations, and the Annenberg study conducted during 2000 US Presidential election provides us with an opportunity to examine a second question-format related issue. In this study, respondents were asked to evaluate the personal characteristics of both Presidential candidates (George W. Bush and Al Gore). Specifically, they were asked for their perceptions of each of the candidates on five traits: “really cares about people like me,” “honest,” “inspiring,” “knowledgeable,” and “strong leadership.”<sup>5</sup> Similarly to the Canadian Election Study in 1993 and 1997, respondents were asked to evaluate the extent to which these traits described the leaders (the options were: extremely well, quite well, not too well, or not well at all). However, the way in which the traits and leaders were presented to respondents varied: half of the sample was asked to consider all traits for one leader before moving on to consider all of the same traits for the other leader (in this study this format is labeled “traits within candidates”), while the other half of the sample was asked to evaluate each leader on a particular trait before moving on to evaluate each leader on the next trait (this format is labeled “candidates within traits”).

While both formats can broadly be seen as comparative (in that by grouping traits or candidates the respondent is considering each trait in comparison to *something*), respondents in each group are effectively comparing different things. In the “traits within candidates” format, respondents are more directly considering a leader’s possession of a trait in comparison with his possession of other traits. In the “candidates within traits” format, the leaders are more obviously juxtaposed, and effectively, respondents are left to compare leaders with each other in the possession of each trait. This randomization exercise provides us with the opportunity to evaluate the effects of these two question formats, and may shed some light on the different types of considerations made by respondents when evaluating leaders.

The analysis of the 2000 Annenberg data was conducted in two steps. First, an analysis of the effects of the randomization on the mean scores for each trait and each leader was conducted,

---

<sup>5</sup> A smaller sample was also asked for their perceptions of two additional traits, but these have been excluded from the overall analysis because they were asked in the post-election period only, and as a result their inclusion in the analysis drastically reduces the overall sample size.

using T-Tests, to see whether respondents were more likely to perceive leaders' traits more or less positively depending on the format that was presented to them. Second, regression analyses were conducted using overall thermometer scores for each leader, mirroring the analyses conducted for the 1993-2000 Canadian Election Studies above, in order to determine whether question format affected overall evaluations of leaders.

What effects might we expect to see? Research on voters' evaluations of leaders suggests that voters evaluate leaders comparatively – that is they compare them to each other, and do not simply evaluate them in a 'black hole' of their own (Brown et al. 1988; Nadeau, Niemi, and Amato 1996). It is conceivable, therefore, that juxtaposing leaders for respondents to evaluate keeps respondents more "honest" in their evaluations of each specific trait. That is, it helps to keep them from over-inflating or over-deflating scores based on a general view of the leader, without thinking about the other leader. It is hypothesized, therefore, that ratings of traits should have a greater impact on thermometer ratings in the "candidates within traits" format, because if a respondent is comparing candidates on a particular trait, and scores a candidate more positively or negatively, this score should be closer to the respondent's "true" perception and should have a stronger impact on overall ratings of the candidate.

Tables 5 and 6 illustrate the results obtained in the initial analysis of the effects of question format on perceptions of leaders' traits. All trait scores were recoded on a 0-1 scale, in which 1 reflected those who felt the trait described the leader "extremely well" while 0 reflected those who felt it did not describe the leader at all. T-Tests were run, comparing the mean scores for each trait for Bush and Gore, by question format. The tests indicate that indeed, question format had an effect (albeit small) on mean trait scores for each of the leaders. With the exception of Gore's perceived honesty, both leaders achieved higher traits scores in the "candidates within traits" format. That is, when the leaders were juxtaposed against each other, they were each seen to be more caring, inspiring, knowledgeable, and to have strong leadership skills than when respondents evaluated all traits for one leader and then moved on to evaluate all traits for the next leader.<sup>6</sup> In contrast, Gore was given higher ratings for "honesty" when this trait was evaluated in conjunction with his other personality traits, rather than in comparison with George Bush's honesty. This exception is interesting, as it was the personality trait which arguably received the most "play" throughout the election campaign, and Gore's honesty was questioned throughout the campaign (Johnston et al. 2004; Geer 2006). It is possible that when respondents compared Gore with Bush on this trait, the difference became more obvious than when they were thinking about Gore's honesty in relation to his knowledge levels or leadership skills. This is the type of consideration which may lead to a stronger relationship between trait and overall thermometer ratings, as hypothesized above.

All of these differences in means were very small—on the order of one to two percent—however, most are statistically significant, as indicated by the size of the standard errors in the two tables.<sup>7</sup> On a most basic level, these differences suggest that question format does have an effect on the responses obtained in the study. Do the differences matter, however? What type of effect do they have on overall evaluations of leaders? The final part of the analysis examines the effects of traits

---

<sup>6</sup> This does suggest that a partisan effect may be taking place here, where Republicans rate Bush more favourably in comparison to Gore, and vice versa. These t-tests do not control for partisanship, thus the regression analysis which follows is essential for teasing out and controlling for these effects.

<sup>7</sup> The two exceptions are "honest" for Bush and "cares" for Gore – these differences did not achieve traditional levels of statistical significance.

and question format on overall ratings of George Bush and Al Gore, as presented in Table 7. This part of the analysis mirrors the regression analyses conducted with the Canadian data, for Alexa McDonough, Gilles Duceppe, and Jean Chretien, in which the thermometer ratings of the leaders were regressed on traits, a question format dummy variable, interactions between traits and question format, and a battery of demographic variables, in addition to party identification.<sup>8</sup>

Table 7 illustrates the effects of traits and question format on overall leader evaluations of both George Bush and Al Gore. As the data indicate, some traits matter more for overall evaluations of Bush, while others matter more for overall evaluations of Gore. Furthermore, the question format has varying effects for each of the leaders. Let us begin by looking at the coefficients for evaluations of George Bush. The first five coefficients reflect the effect of traits on overall evaluations among respondents who were presented with traits in the “candidates within traits” format—these individuals were asked to evaluate a trait for each of the leaders before moving on to the next trait. These data suggest that those who felt that “cares” described Bush “extremely well” tended to give him overall thermometer scores nearly 16 points higher on a 0-100 scale than those who felt that the trait did not describe Bush at all. In comparison, for those evaluating all of Bush’s traits before moving on to Gore’s traits, those individual who felt that he “cares” gave him overall thermometer ratings 15 points higher (coefficient 0.151; standard error 0.010) than those who felt that the trait did not describe him – a decrease of one point on the 0-100 scale.

Individuals giving the highest rating to Bush for the traits honest and inspiring when compared to Gore’s level of honesty and inspiration were likely to rate him 15 points higher on the 0-100 scale than those who did not feel he was honest or inspiring. This changed slightly for those evaluating Bush in the other question format, when all traits were evaluated for Bush before moving on to Gore, with statistically significant coefficients of 0.17 and 0.158, respectively. This illustrates an increase in the impact of these traits by 1-2 points on the 0-100 thermometer scale. Thus for these two traits, Bush’s overall evaluations were more positively affected when his possession of these traits was compared with his possession of all other traits.

Whether or not Bush was perceived to be knowledgeable seems to have mattered less in comparison to the other traits for overall feelings about him—those who felt that knowledgeable described him “extremely well” rated him 10 points higher on the thermometer scale than those who did not think he was knowledgeable. This trait was even less important for those comparing Bush to Gore on this trait, as the impact of a positive evaluation on this trait led to an 8 point increase (coefficient 0.079; standard error 0.01) in overall thermometer ratings of Bush (2 points lower than among those receiving the first question format—his overall ratings were more strongly affected by those comparing him to Gore on this trait).

It was the issue of his leadership that had the greatest influence on his overall evaluations, with those giving him the highest ratings for “strong leadership” gave him thermometer ratings nearly 22 points higher than those who did not feel that this trait described Bush at all. In the other question format, those feeling that strong leadership described Bush “a great deal” gave him thermometer ratings nearly 18 points higher than those who did not feel he possessed strong leadership skills. Thus this trait was most important in explaining overall thermometer ratings (a four point difference) when the two leaders were directly juxtaposed.

---

<sup>8</sup> Again, see Appendix B for details of coding of all variables.

For evaluations of Al Gore, the effects of question format were much more substantial, although not all interactive coefficients achieved traditional levels of statistical significance. Among those individuals evaluating Gore's personality traits within the "candidates within traits" format, those who felt he "cares" gave him ratings nearly 19 points higher than those who felt that this trait did not describe him at all. The impact of this trait on overall evaluations was less strong for those individuals comparing the extent to which he cares with the extent to which Gore possessed all of the other traits: among those receiving the second question format, those who felt that "cares" described Gore "a great deal" gave him overall thermometer ratings just over 10 points higher than those who felt that this trait did not describe Gore at all (coefficient 0.101; standard error 0.033).

The next three traits did not achieve traditional levels of statistical significance in the "traits within candidates format," thus while those who felt that he was honest, inspiring, and knowledgeable gave him higher thermometer ratings than those who didn't feel he possessed these traits by 20 points, 12 points, and 9 points, respectively, we cannot really compare these coefficients to those in the other question format (among those receiving the traits within candidates format, the coefficient for honest was 0.015; s.e. of 0.652; the coefficient for inspiring was 0.047; s.e. of 0.180; and the coefficient for knowledgeable was -0.027; s.e. of 0.031).

The impact of perceptions of his leadership skills, however, can be compared by question format. Among those comparing his leadership skills to Bush, respondents who felt that "strong leadership" described Gore "extremely well" gave him overall thermometer scores 23 points higher on the 0-100 scale than those who did not feel that the trait described him at all. In contrast, among those who evaluated the strength of Gore's leadership in conjunction with his possession of the other traits, the impact of this trait was substantially decreased. Individuals receiving this question format gave Gore overall thermometer ratings 13 points higher when they felt that "strong leadership" described him "extremely well" compared to those who did not feel that it described him at all – a difference of 10 points from one question format to the other. Thus for Gore, positive evaluations of his personality traits when made side by side with evaluations of Bush's personality traits had a much greater impact on his overall evaluations than did evaluations of his traits when made in comparison to his other traits.

The hypothesis that positive trait evaluations would have a greater impact on overall evaluations if the traits were evaluated in juxtaposing the two leaders against each other was therefore partially supported by these data. For Bush, evaluations of the extent to which he "cares," is "knowledgeable," and possesses "strong leadership" skills had a greater impact on overall thermometer ratings when these traits were evaluated side by side with the extent to which Gore possesses these traits. For Gore, the extent to which he "cares," and possesses "strong leadership" skills had a greater impact on overall evaluations when these traits were compared with the extent to which Bush possessed the same traits. However, for Bush, "honest" and "inspiring" had a greater impact on overall evaluations when they were evaluated in conjunction with his other traits, and for Gore, three of the interactive coefficients did not achieve traditional levels of statistical significance, leaving us unable to draw any conclusions about them.

These data do suggest that broadly, the way in which questions are asked does indeed have an impact on the responses we get. Why it is that the effect of evaluations leaders' strength of leadership most varied by question format is not entirely known. It may have something to do with the level of attention this trait received during the campaign—perhaps this was the "important" trait, and question format stimulated different thought processes during the

evaluation of this trait. Johnston et al. note that doubts about Bush's leadership qualities grew as the campaign progressed (2004:127), an observation which may help to explain why this trait had the largest impact on overall evaluations regardless of question format. That the impact of the trait was largest when the leaders were juxtaposed in evaluating it fits with the notion that voters do not evaluate candidates in a "black hole" of their own, but that evaluations are performed on a comparative basis. The results, however, are inconclusive, and more research into the effects of question format on trait evaluations, and how these effects translate onto overall leader evaluations—and even vote choice—is certainly needed.

## Discussion

In a study of the effects of leadership in Britain and Australia, Bean and Mughan (1989) examined all nine of the traits commonly probed in both the Australian and British elections of 1987 and 1983, respectively. They found that the same four characteristics were important in both countries: *effectiveness, listening to reason, caring, and sticking to principles*. A few years later, in a study of Australia and New Zealand, Bean (1993) grouped open-ended evaluations of leaders into seven main categories: *competence, integrity, strength, harmony, general likeability, other personal, and policy/party/group*, and found that *competence* and *integrity* were the qualities taken into account most by voters when making decisions. The traits focused on in each study reflect the data available. These two studies illustrate one of the main difficulties with comparative research into the importance of leadership traits: the comparability and transportability of the survey questions themselves.

A number of studies (see, for example, Bean 1993; Glass 1985; Kinder et al. 1980; Miller et al. 1986; Brettschneider and Gabriel 2002) are based on open-ended questions about leaders, such as the American NES standard (this example taken from 2004 NES):

*Qa3 - Now I'd like to ask you about the good and bad points of the major candidates for President. Is there anything in particular about George W. Bush that might make you want to vote for him? What is that?*

*Qa4 - Is there anything in particular about George W. Bush that might make you want to vote AGAINST him? What is that?*

Meanwhile, numerous others are based on closed-ended questions (Bean and Mughan 1989; Brown et al. 1988; Johnston 2002; Johnston et al. 2004). There are substantial differences between these two question types, limiting the extent of their comparability, but even within closed-ended traits questions, substantial differences in question format exist, as outlined earlier, which may restrict our ability to conduct large scale comparative research into the importance of leader traits, thus limiting how much we can really "know" about how the evaluation of party leaders matters in election outcomes.

This study is an attempt to begin to unpack the effects of question format on how voters evaluate leaders' personality traits, and clearly, more research is needed. The results of the analysis are somewhat mixed, for both the Canadian and US analyses. What is clear is that question format does have an effect, although the effects seen were not always large, and did not necessarily conform to expectations. One of the issues with the Canadian data is the existence of only the one negative trait in the three election studies in question, making it hard to determine whether the

patterns which emerged were related to negative versus positive traits, or whether they were just related to the one negative trait, the leaders' arrogance. Another challenge with the Canadian data is the difficulty of separating out question format effects from "election year" effects. Extending the analysis to include 1993 CES data was a partial solution, but even this isn't as reliable as an analysis based on data in which these types of issues do not arise, as in a single election study with different question formats present (like the 2000 Annenberg study). The other problem is that over time, leaders change, and so comparing trait evaluation and question effects from year to year is difficult. Some thought needs to be given to furthering research design to take these issues into consideration, before we can determine with more certainty what effects question format really have on leader evaluations.

The other question that we will only really be able to answer with further research is the "does this really matter" question. This study has focused on the effects of question format on trait evaluations, and in particular, the effects of these trait evaluations on overall leader evaluations. The results have been mixed, and while they are largely statistically significant, many of the results are small, and the degree to which they are substantively significant needs to be addressed further. For example, what type of effect does question format have on overall electoral results? Future research into the role of question format must take the analysis one step further, and consider the "ultimate" impact of these issues.

The issue of measurement does matter, particularly when we consider the methodological and conceptual difficulties of large-scale comparative research into the question of the role of voters' evaluations of leaders. Much of the literature remains divided as to the issue of whether leaders matter, how much they matter, but especially as to which traits really matter. These kinds of divisions, this paper argues, are partially a result of the data upon which conclusions are based, and the lack of continuity of these data across time and space. In order to understand the role of leaders' personality traits in electoral outcomes, a broad, comparative study is crucial. In order to be able to conduct such a study, a number of methodological kinks must be worked out. This paper can be seen as an attempt to begin to look into one of these kinks, the effect of question format on the evaluation of closed-ended questions surrounding leaders' personality traits. The results suggest that this is an issue which requires further exploration.

## Appendix A: Tables

Table 1 Closed-Ended Traits Included in Surveys Across Space and Time		
Root Trait	Traits	Countries/years
	able to stand up to Quebec demands	Canada 1968
	adventurer (daring)/balanced and sensible	Israel 1981
	aggressive	Canada 1993
		Australia 1996, Canada 1968, Canada 1984, Canada 1993, Canada 1997, Canada 2000, Germany 1980, Germany 1987, NZ 1999, NZ 2002, Annenberg 2004
	arrogant	Canada 1993, NZ 1999
	can really speak for women	
	capable of solving English-French problems in Canada	Canada 1968
		Australia 1987, Britain 1983, Britain 1987, Britain 1992
	caring	Canada 1968
	charming	Canada 1984, US 1984
	commands respect	Canada 1968
	pro-communist	Canada 1968
		Australia 1993, Australia 1996, Australia 1998, Australia 2001, Australia 2004, Canada 1988, Canada 1993, Canada 1997, Canada 2000, Netherlands 1983, NZ 1999, US 1984, US 1988, US 1992, US 1996
competent	compassionate	Canada 2004, Canada 2006
	competent	Britain 2005, Canada 1984
	competent leader	Netherlands 1983
	willing to compromise	Germany 1980, Germany 1987
conservative	conservative/modern	Canada 1968
	views are too conservative	Annenberg 2004
	too conservative	
	credible in TV appearances/not credible in TV appearances	Israel 1981
	would control crime	US 1972
		Australia 1993, Canada 1984, US 1984, US 1988
	decent	Australia 1987, Britain 1983, Britain 1997, Israel 1981, Israel 1992, Israel 1996, Israel 1999, Netherlands 1983, Annenberg 2004
decisive	decisive	US 2004
	can't make up his mind	Australia 1993, Australia 1996, Australia 1998
	dependable	Australia 1987, Britain 1983
	determined	Germany 1998
	diligent	Canada 1984
	dull	Germany 1980, Germany 1987
	dynamic/hesitating	

extreme	experienced	Israel 1996, 1999
	extreme	Canada 2000
fair	extreme/moderate	Britain 1987, Britain 1992
	Stephen Harper is just too extreme	Canada 2004, Canada 2006
fresh	fair	US 1984
	fair-minded	Canada 1968
gets things done	tired/fresh	Israel 1981
	young leader/old leader	Israel 1981
	a leader whose time has passed/a leader whose time has not passed	Israel 1981
	his health does not permit him to continue in his role/his health permits him to continue in his role	Israel 1981
	represents change	Canada 1984
	having new ideas	Canada 2000
	gets things done	US 1992, US 1996
	good at getting things done/bad at getting things done	Britain 1987
	effective & gets things done	Annenberg 2004
	initiates and moves things	Israel 1992, Israel 1996, Israel 1999
honest	hardworking	US 1984
	is helpful to his country/is damaging to the country	Israel 1981
	honest	Australia 1996, Australia 1998, Australia 2001, Australia 2004, Canada 1968, Canada 2004, Canada 2006, Israel 1981, US 1988, US 1992, US 1996, Annenberg 2000, Annenberg 2004
	is an honest and trustworthy person	Russia 1995/96
	dishonest	Canada 2000, US 1980, US 2000, US 2004
	hypocritical	Annenberg 2000
	ill-tempered, lacking control/calm, matter of fact	Germany 1980, Germany 1987
	inspiring	Australia 1993, Australia 1996, Australia 1998, Australia 2001, Australia 2004, Sweden 1988, Sweden 1991, US 1980, US 1984, US 1988, US 1992, US 1996, Annenberg 2000, Annenberg 2004
	in touch with the times	Canada 1997
	intelligent	Australia 1993, Australia 1996, Australia 1998, Australia 2001, Australia 2004, Canada 1984, Canada 1993, Canada 2000, US 1984, US 1988, US 1992, US 1996, US 2000, US 2004
highly intelligent	Canada 1968	
is an intelligent and knowledgeable	Russia 1995/96	

	person	Canada 1968
	man of great integrity	US 1984
	kind	Australia 1993, Australia 1996, Australia 1998, Australia 2001, Australia 2004, Canada 1988, Sweden 1988, Sweden 1991, US 1980, US 1984, US 1988, US 1992, US 1996, US 2000, US 2004, Annenberg 2000, Annenberg 2004
	knowledgeable	Annenberg 2004
	too liberal	Annenberg 2004
likeable	likeable as a person	Australia 1987, Britain 1983, Britain 1987
	he's easy to like as a person	Annenberg 2004
	listens to reason	Australia 1987, Britain 1983
	given to moods/not given to moods	Israel 1981 Australia 1993, Australia 1996, Australia 1998, Canada 1988, US 1980, US 1984, US 1988, US 1992, US 1996, US 2000, US 2004
moral	moral	
	which candidate best reflects high moral or religious standards	US 1972
	would bring moral and religious standards to government	US 1976
	has adequate concern for public moral standards	Canada 1968 Israel 1992, Israel 1996, Israel 1999, Israel 2003
	good negotiator	Canada 1984
	nervous	Annenberg 2004
	optimistic	Israel 1992, Israel 1996, Israel 1999, Israel 2001, Israel 2003
power- hungry	places benefit of country before the party Paul Martin only cares about staying in power	Canada 2006 US 1980
	power-hungry	Canada 1968
	progressive	
keeps promises	someone who keeps his promises/breaks his promises	Britain 2001 Canada 1968
	too quick to make promises	Annenberg 2004
	reckless	Australia 1993, Australia 1996, Australia 1998, Germany 1998, Israel 1992, Israel 1996, Israel 1999, Israel 2001, Israel 2003, Netherlands 1983, Sweden 1988, Sweden 1991
	reliable	US 1984
	religious	Germany 1980, Germany 1987, Germany 1998
responsive to 'ordinary'	responsible/irresponsible	
	shares my values	Annenberg 2004

## values

	listens to the views of people in (name province)	Canada 1984
	Paul Martin only cares about big business	Canada 2004
	Jack Layton only cares about minorities	Canada 2004
	in touch with ordinary people	US 1984
	looks after one class/looks after all classes	Britain 1987, Britain 1992
	understands people like you	US 1984
	out of touch with ordinary people	US 2000
	out of touch with people like me	Annenberg 2004
	cares about people like me	Annenberg 2004
		Canada 1988, Russia 1995, US 1984, US 1988, US 1992, US 1996, US 2000, US 2004
	really cares about people like you	
	knows the thoughts and opinions of ordinary people	Sweden 1988, Sweden 1991
	responds to voters concerns	Britain 2005
	too rigid	Canada 1968
'right' leadership skills	better suited for provincial than for federal politics	Canada 1968
	a bad Prime Minister/an excellent Prime Minister	Israel 1981
	a bad Defense Minister/an excellent Defense Minister	Israel 1981
	has the kind of personality a President ought to have	US 1972, US 1976
	has the right kind of experience to be President	Annenberg 2004
	his place is in the opposition/his place is in government	Israel 1981
	ruthless	Canada 1984
		Australia 1993, Australia 1996, Australia 1998, Australia 2001, Australia 2004
	sensible	US 1984
	sets a good example	Canada 1984
	shallow	Australia 1987, Britain 1983
	shrewd	Australia 1987, Canada 1984
	sincere	Israel 1981
slick	slick/straight	Canada 1984
	slick	Canada 1968
	manner is too slow	Canada 1968
	too soft on French Canada	Canada 1968
	stable/unstable	Israel 1981
	steady	Annenberg 2004
	can stand stress/cannot stand stress	Israel 1981
sticks to principles	sticks to principles	Australia 1987, Britain 1983, Britain 1997
	says one thing but does another	Annenberg 2004
	someone who will make the tough decisions despite political pressure	Annenberg 2004

	can stand up under pressure	Israel 1992, Israel 1996, Israel 1999, Israel 2001
	changes his mind for political reasons	Annenberg 2004
	flip-flops on the issues	Annenberg 2004
	not willing to admit when he makes a mistake	Annenberg 2004
strong leader	strong leader	Canada 1997, Russia 1995, Annen 2004
		Australia 1993, Australia 1996, Australia 1998, Australia 2001, Australia 2004, Canada 1988, Canada 1993, NZ 1999, NZ 2002, US 1984, US 1988, US 1992, US 1996, US 2000, US 2004,
	provides strong leadership	Annenberg 2000
	strong/weak	Israel 1981, Israel 1992
	has a leadership characteristic	Israel 1992, Israel 1996, Israel 1999, Israel 2003
	capable of being a strong leader/not capable of being a strong leader	Britain 1987, Britain 1992, Britain 1997, Britain 2001
	a weak leader	Canada 2000
	weak	US 1980
	stubborn	Annenberg 2004
	sympathetic	Germany 1998
	sure of himself	Canada 1984
	can work on a team/cannot work on a team	Israel 1981
tough	tough	Australia 1987, Britain 1983
	tough-minded	Canada 1968
		Australia 2001, Australia 2004, Britain 2005, Canada 1988, Canada 1993, Canada 1997, Canada 2000, Germany 1980, Germany 1987, NZ 1999, NZ 2002, Annenberg 2000,
trustworthy	trustworthy	Annenberg 2004
	could be trusted	US 1972, US 1976
	Jack Layton cannot be trusted	Canada 2006
	Gilles Duceppe cannot be trusted	Canada 2006
	likely to unite the nation/divide the nation	Britain 1987
	he has a clear vision of where he wants to lead the country	Annenberg 2004
has vision	has his own vision of the country's future	Russia 1995/96
	man of vision	Canada 1988
	under Dalton Camp's influence	Canada 1968
	warm	Canada 1984
	weighs his words carefully/does not weigh his words carefully	Israel 1981
	will know how to fight terror	Israel 2001
	will negotiate decisively in the negotiation on peace and terror	Israel 2001
	will lead to real peace with Arabs	Israel 2001
	will reduce the division within Israeli	Israel 2001

society  
would reduce the tension between  
religious and seculars

Israel 1999  
Israel 1999, Israel 2001, Israel  
2003

would preserve the rule of law  
would reduce social gaps  
would bring peace in Vietnam

Israel 1999  
US 1972

---

	1968	1984	1988	1993	1997	2000	2004	2006
Question format	agree/disagree	scale of 1-7	level of agreement: strongly/somewhat agree/disagree			choose a leader	scale of 0-10	
Trait	integrity intelligent honest	arrogant competent ruthless commands respect dull warm nervous decent slick sincere shallow sure of self	intelligent         trustworthy compassionate knowledgeable moral strong leader	intelligent  arrogant    trustworthy compassionate strong leader	arrogant    trustworthy compassionate strong leader	intelligent  arrogant   trustworthy compassionate weak leader	arrogant	honest  competent      can't be trusted
Party Leaders	Trudeau Stanfield	Turner Mulroney Broadbent	Turner Mulroney Broadbent	Chretien Campbell McLaughlin Bouchard Manning	Chretien Charest McDonough Duceppe Manning	Chretien Clarke McDonough Duceppe Day	Martin Harper Layton Duceppe	Martin Harper Layton Duceppe

Table 3:  
Effects of Traits and Question Format on Feelings Towards Alexa McDonough and Gilles Duceppe

	McDonough	Duceppe
Trait: Arrogant	-0.129 (0.067)	-0.136 (0.030)**
Trait: Trustworthy	0.144 (0.017)**	0.150 (0.024)**
Trait: Compassionate	0.076 (0.012)**	0.061 (0.025)*
Question Format (extent of agreement)	-0.206 (0.030)**	-0.250 (0.033)**
Arrogant*qformat	0.028 (0.072)	0.083 (0.043)
Trustworthy*qformat	0.193 (0.039)**	0.259 (0.045)**
Compassionate*qformat	0.026 (0.038)	-0.018 (0.048)
Age	0.000 (0.000)	-0.001 (0.000)*
Woman	0.046 (0.009)**	0.013 (0.014)
Degree	0.030 (0.010)**	0.014 (0.015)
Income	0.007 (0.006)	-0.004 (0.010)
Non-religious	0.004 (0.013)	-0.009 (0.026)
Visible Minority	0.023 (0.024)	-0.005 (0.039)
Atlantic	0.035 (0.015)*	
Quebec	-0.039 (0.015)**	
West	-0.000 (0.012)	
Liberal PID	0.025 (0.012)*	-0.044 (0.020)*
Conservative PID	0.010 (0.016)	0.043 (0.035)
NDP PID	0.143 (0.018)**	0.055 (0.055)
Reform/Alliance PID	-0.039 (0.017)*	-0.097 (0.069)
Bloc Quebecois PID	-0.024 (0.020)	0.143 (0.019)**
Constant	0.342 (0.026)**	0.491 (0.039)**
Observations	2172	912
R-Squared	0.27	0.43

Standard errors in parentheses

\* significant at 5% level; \*\* significant at 1% level

Table 4: Effects of Trait and Question Format on Feelings Towards Jean Chretien

Trait: Arrogant	-0.170 (0.011)**
Trait: Trustworthy	0.168 (0.015)**
Trait: Compassionate	0.075 (0.016)**
Question Format (extent of agreement)	-0.204 (0.015)**
Arrogant*qformat	0.093 (0.015)**
Trustworthy*qformat	0.134 (0.020)**
Compassionate*qformat	0.069 (0.021)**
1997	-0.001 (0.007)
Age	-0.001 (0.000)**
Woman	0.000 (0.006)
Degree	0.031 (0.006)**
Income	0.008 (0.004)*
Non-religious	-0.009 (0.008)
Visible Minority	0.026 (0.012)*
Atlantic	-0.008 (0.010)
Quebec	-0.056 (0.008)**
West	-0.012 (0.007)
Liberal PID	0.124 (0.007)**
Conservative PID	0.017 (0.009)*
NDP PID	0.043 (0.011)**
Reform/Alliance PID	-0.045 (0.011)**
Constant	0.516 (0.017)**
Observations	5386
R-Squared	0.43

Standard errors in parentheses

\* significant at 5% level; \*\* significant at 1% level

Table 5: Effects of Randomization on Bush's Trait Scores

	Candidates within Traits	Traits within Candidates	Difference
Honest	0.553 (0.002)	0.555 (0.001)	-0.002 (0.003)
Cares	0.486 (0.002)	0.473 (0.002)	0.014 (0.003)
Inspiring	0.478 (0.002)	0.457 (0.002)	0.021 (0.002)
Knowledgeable	0.578 (0.002)	0.562 (0.002)	0.015 (0.003)
Leadership	0.567 (0.002)	0.56 (0.002)	0.007 (0.003)

Table 6: Effects of Randomization on Gore's Trait Scores

	Candidates within Traits	Traits within Candidates	Difference
Honest	0.496 (0.002)	0.503 (0.002)	-0.007 (0.003)
Cares	0.511 (0.002)	0.508 (0.002)	0.003 (0.003)
Inspiring	0.427 (0.002)	0.414 (0.002)	0.013 (0.002)
Knowledgeable	0.661 (0.002)	0.645 (0.002)	0.017 (0.003)
Leadership	0.534 (0.003)	0.519 (0.003)	0.015 (0.004)

Table 7: Effects of Traits and Randomization on Feelings Towards George Bush and Al Gore		
	Bush	Gore
Trait: Cares	0.157 (0.010)**	0.189 (0.010)**
Trait: Honest	0.151 (0.010)**	0.198 (0.010)**
Trait: Inspiring	0.152 (0.010)**	0.118 (0.010)**
Trait: Knowledgeable	0.098 (0.010)**	0.087 (0.010)**
Trait: Leadership	0.217 (0.011)**	0.231 (0.011)**
Randomization (Traits within Candidates)	0.006 (0.008)	0.001 (0.008)
Cares*Randomization	0.005 (0.014)	-0.022 (0.014)
Honest*Randomization	0.029 (0.014)*	0.047 (0.014)**
Inspiring*Randomization	0.015 (0.014)	0.007 (0.014)
Knowledgeable*Randomization	-0.016 (0.014)	0.007 (0.014)
Leadership*Randomization	-0.030 (0.015)	-0.029 (0.015)
Republican PID	0.080 (0.004)**	-0.088 (0.004)**
Democrat PID	-0.086 (0.004)**	0.072 (0.004)**
Woman	-0.008 (0.003)*	0.014 (0.003)**
Black	-0.059 (0.006)**	0.038 (0.005)**
Hispanic	0.005 (0.006)	0.037 (0.006)**
Degree holder	0.012 (0.003)**	0.027 (0.003)**
Married	0.002 (0.003)	-0.015 (0.003)**
Employed	0.006 (0.004)	0.000 (0.003)
Evangelical Christian	0.019 (0.003)**	-0.031 (0.003)**
Income	0.006 (0.001)**	0.003 (0.001)**
Union	-0.012 (0.004)**	0.013 (0.004)**
Constant	0.130 (0.008)**	0.099 (0.008)**
Observations	20214	20404
R-squared	0.56	0.61

Standard errors in parentheses

\* significant at 5% level; \*\* significant at 1% level

## Appendix B: Construction of Variables

### *Canadian Election Studies*

Leader Thermometers: all recoded on a 0-1 scale (although effects of coefficients are reported in the text as if on a 0-100 scale); don't know recoded to mid-point (0.5); refused recoded as missing.

Traits: all trait evaluations recoded on a 0-1 scale, DK/Ref coded as missing. In 2000, all those selecting a leader for a trait or "all leaders" recoded as 1; leader not chosen or respondent chooses "none" recoded as 0.

Question format variable: dummy variable constructed; all those who were presented with the 1993 and 1997 traits questions (categories of agreement) coded as 1; all those presented with the 2000 "choice" format coded as 0.

Interactions created by multiplying each trait for each leader by question format dummy variable (i.e. Chretien $\text{trait} \times \text{format}$ ; McDonough $\text{trait} \times \text{format}$ ; Duceppe $\text{trait} \times \text{format}$ )

Woman: dummy variable constructed, 1 = woman, 0 = man

Degree holder: education variable recoded into dummy variable, in which those holding a college or university degree are coded as 1; all others = 0. DK/Ref = missing.

Income: recoded into three categories-respondents divided into bottom, middle, and top thirds of income levels in each election year. DK/Ref = missing.

Non-religious: dummy variable created, all those claiming no religion = 1; those with any religious affiliation = 0. DK/Ref = missing.

Visible Minority: based on "ethnicity" questions in the CES; all those of visible minority coded 1, others = 0. DK/Ref = missing.

Atlantic/West: coded based on "province" variable; respondents from any Atlantic province coded 1, all others 0; respondents from Manitoba-BC coded as 1, all others 0.

PID variables: dummy variables created from partisanship variables in CES; those claiming partisanship with one of the parties = 1, all others = 0. DK/Ref = missing.

### *National Annenberg Election Study*

Leader Thermometers: all recoded on a 0-1 scale (although effects of coefficients are reported in the text as if on a 0-100 scale); don't know recoded to mid-point (0.5); refused recoded as missing.

Traits: all trait evaluations recoded on a 0-1 scale, DK/Ref coded as missing.

Randomization: based on variable `cazo1`, recoded into dummy variable, in which those respondents presented with the “traits in candidates” format coded as 1, and “candidates in traits” coded as 0.

Interactions created by multiplying each trait for each leader by question format dummy variable (i.e. `Bushtrait*format`; `Goretrait*format`)

PID variables: dummy variables created from partisanship variables in Annenberg election study, those claiming partisanship with one of the parties = 1, all others = 0. DK/Ref = missing.

Woman: dummy variable constructed, 1 = woman, 0 = man.

Black: dummy variable constructed, 1 = black, 0 = other.

Hispanic: dummy variable constructed, 1 = Hispanic, 0 = other.

Degree holder: education variable recoded into dummy variable, in which those holding a college or university degree are coded as 1; all others = 0. DK/Ref = missing.

Married: dummy variable constructed, 1 = married/living with partner, 0 = other. DK/Ref = missing.

Employed: dummy variable constructed, 1 = employed, 0 = unemployed. DK/Ref = missing.

Evangelical Christian: dummy variable constructed, 1 = Evangelical Christian, 0 = other religious affiliation/no affiliation. DK/Ref = missing.

Income: recoded into three categories-respondents divided into bottom, middle, and top thirds of income levels in each election year. DK/Ref = missing.

Union: dummy variable constructed, 1 = union members in household, 0 = non-union. DK/Ref = missing.

## References

- Banducci, Susan A., and Jeffrey A. Karp. 2000. Gender, Leadership and Choice in Multiparty Systems. *Political Research Quarterly* 53 (4):815-848.
- Bean, Clive. 1993. The Electoral Influence of Party Leader Images in Australia and New Zealand. *Comparative Political Studies* 26:111-32.
- Bean, Clive, and Anthony Mughan. 1989. Leadership Effects in Parliamentary Elections in Australia and Britain. *American Political Science Review* 83:1165-79.
- Brettschneider, Frank, and Oscar W. Gabriel. 2002. The Nonpersonalization of Voting Behavior in Germany. In *Leaders' Personalities and the Outcomes of Democratic Elections*, edited by A. King. Oxford: Oxford University Press.
- Brown, Steven D., Ronald D. Lambert, Barry J. Kay, and James E. Curtis. 1988. In the Eye of the Beholder: Leader Images in Canada. *Canadian Journal of Political Science* 21:729-55.
- Clarke, Harold D., Karl Ho, and Marianne C. Stewart. 2000. Major's Lesser (Not Minor) Effects: Prime Ministerial Approval and Governing Party Support in Britain Since 1979. *Electoral Studies* 19:255-73.
- Conover, Pamela J., and Stanley Feldman. 1989. Candidate Perception in an Ambiguous World: Campaigns, Cues, and Inference Processes. *American Journal of Political Science* 33 (4):912-940.
- Crewe, Ivor, and Anthony King. 1994. Did Major Win? Did Kinnock Lose? Leadership Effects in the 1992 Election. In *Labour's Last Chance? The 1992 Election and Beyond*, edited by A. Heath, R. Jowell and J. Curtice. Aldershot: Dartmouth.
- Gant, Michael M., and Lee Sigelman. 1985. Anti-Candidate Voting in Presidential Elections. *Polity* 18 (2):329-339.
- Geer, John. 2006. *In Defense of Negativity: Attack Ads in Presidential Campaigns*. Chicago: University of Chicago Press.
- Gidengil, Elisabeth, André Blais, Neil Nevitte, and Richard Nadeau. 2000. Do Election Campaigns Prime Leadership? Evidence From Recent Canadian Elections. Paper read at ECPR 28th Joint Session Workshops, at Copenhagen.
- Glass, David P. 1985. Evaluating Presidential Candidates: Who Focuses on Their Personal Attributes? *The Public Opinion Quarterly* 49 (4):517-534.
- Graetz, Brian, and Ian McAllister. 1987a. Party Leaders and Election Outcomes in Britain, 1974-1983. *Comparative Political Studies* 19:484-507.

- Graetz, Brian, and Ian McAllister. 1987b. Popular Evaluations of Party Leaders in the Anglo-American Democracies. In *Political Elites in Anglo-American Democracies*, edited by H. D. Clarke and M. M. Czudnowski. DeKalb, Illinois: Northern Illinois University Press.
- Hudson, John. 1984. Prime Ministerial Popularity in the UK: 1960-81. *Political Studies* 32:86-97.
- Johnston, Richard. 2002. Prime Ministerial Contenders in Canada. In *Leaders' Personalities and the Outcomes of Democratic Elections*, edited by A. King. Oxford: Oxford University Press.
- Johnston, Richard, Michael G. Hagen, and Kathleen Hall Jamieson. 2004. *The 2000 Presidential Election and the Foundations of Party Politics*. Cambridge: Cambridge University Press.
- Kinder, Donald R., Mark D. Peters, Robert P. Abelson, and Susan T. Fiske. 1980. Presidential Prototypes. *Political Behavior* 2:315-37.
- Lewis, Justin, and Karin Wahl-Jorgensen. 2004. The Media Representation of Public Opinion: British Television News Coverage of the 2001 General Election. *Media, Culture, and Society* 26 (1): 63-80.
- Mendelsohn, Matthew. 1994. The Media's Persuasive Effects: The Priming of Leadership in the 1988 Canadian Election. *Canadian Journal of Political Science* 27:81-97.
- Mendelsohn, Matthew. 1996. The Media and Interpersonal Communications: The Priming of Issues, Leaders, and Party Identification. *The Journal of Politics* 58 (1):112-25.
- Miller, Arthur H., and Martin P. Wattenberg. 1985. Throwing the Rascals Out: Policy and Performance Evaluations of Presidential Candidates, 1952-1980. *American Political Science Review* 79:359-72.
- Miller, Arthur H., Martin P. Wattenberg, and Oksana Malanchuk. 1986. Schematic Assessment of Presidential Candidates. *American Political Science Review* 80:521-40.
- Mutz, Diana C. 1995. Effects of Horse-Race Coverage on Campaign Coffers: Strategic Contributing in Presidential Primaries. *The Journal of Politics* 57 (4): 1015-1042.
- Nadeau, Richard, Richard G. Niemi, and Timothy Amato. 1996. Prospective and Comparative or Retrospective and Individual? Party Leaders and Party Support in Great Britain. *British Journal of Political Science* 26 (2):245-258.
- Nakai, Emily. 2003. *Voters' evaluations of prime ministerial candidates: the impact of leader traits in the 2000 Canadian federal election*. Unpublished Masters Thesis: Department of Political Science, University of British Columbia.
- Stewart, Marianne C., and Harold D. Clarke. 1992. The (Un)Importance of Party Leaders: Leader Images and Party Choice in the 1987 British Election. *Journal of Politics* 54:447-70.

Stokes, Donald E., Angus Campbell, and Warren E. Miller. 1958. Components of Electoral Decision. *American Political Science Review* 52:367-87.