Re-evaluating the Valence Model of Political Choice*

GEOFFREY EVANS AND KAT CHZHEN

The influential valence model of voting developed over the last decade by the British Election Study (BES) team assumes that party and leadership performance evaluations have a causal impact on party choice. An alternative perspective argues that such performance evaluations are instead the consequences of party choice. This article examines the analytical and empirical underpinnings of the BES valence model and compares it to the party-driven approach. To do so, it estimates cross-lagged structural equation models of the association between Labour Party preference and evaluations of the Labour government’s performance during the 2005–10 British electoral cycle. It shows that party preference has a stronger effect on performance evaluations than vice versa; performance evaluations have no significant effect on party preference toward the end of the electoral cycle. The study also finds that, contrary to claims made concerning their merits as simplifying heuristics, performance assessments have no impact on short-term movements in party choice for less politically attentive voters. To a substantial degree, evaluations of party performance express—rather than explain—party choice, and would appear to have limited merit as simplifying heuristics.

Stokes (1963, 1994) famously argued that competition between parties involves assessing policy performance and party competence in the delivery of agreed-upon goals. There is assumed to be a general consensus on the desirability of goals such as economic growth, the effective delivery of public services and the reduction of crime. Vote choice is based on which party or candidate is perceived to be the most competent to deliver these sorts of outcomes. This theory contrasts markedly with approaches to voting that emphasize differences in issues and/or ideological positions between parties and voters. In recent years, perceptions of party competence have become important explanations of political choices and electoral outcomes. Voters evaluate parties and politicians in terms of their competence and performance, and vote for those they think have delivered, or will deliver, valued outcomes.1 The significance of evaluating performance and competence for vote choice is now so embedded in our understanding of voting behavior that voter assessments of party competence on key issues are routinely included in election surveys. The American, British, Canadian, European and Irish election studies all now include questions on party issue competence (Wagner and Zeglovits forthcoming).

A particular model of valence voting advocated by the current British Election Study (BES) team has recently become highly influential. Clarke et al. (2004, 2009, 2011) have instantiated an operationalization of valence voting that has reshaped understanding of British voting behavior in particular, but has also been assumed to have more general

* Geoffrey Evans is Official Fellow in Politics, Nuffield College, and Professor in the Department of Politics and International Relations, University of Oxford (geoffrey.evans@nuffield.ox.ac.uk). Kat Chzhen is a Post-doctoral Research Fellow at Nuffield College, and the Department of Politics and International Relations, University of Oxford (yekaterina.chzhen@politics.ox.ac.uk).

1 For example, Fiorina 1981; Fournier et al. 2003; Green and Jennings 2012a,b.
temporal and spatial applications.\(^2\) The BES valence model has been influential because of its thorough and systematic empirical analyses and explicit specification and operationalization of what is involved in valence voting and how it should be measured. One recent reviewer concludes, “the valence model, and the sophisticated ways in which it is tested, provides a powerful, possibly unassailable, explanatory account of the overall pattern of party choice” (Johnston 2011, 213). Another offers that “Clarke et al. (2004) make a compelling case for the predominance of valence voting in recent (and probably not so recent) elections” (Johns 2010, 156). In this article we question whether the valence model advocated by Clarke et al. does actually present an unassailable explanatory account of party choice by examining whether the model of valence/performance voting they advance is empirically valid.

Our primary focus is on whether the key elements of the BES valence model—(1) which is the best performing party on the most important issue, (2) which party is evaluated most highly on salient issues and (3) which is the best party leader—can explain vote intention or reported vote. More specifically, we are interested in whether these associations between answers to questions, often measured sequentially in the same survey, can be treated as causal and unidirectional rather than simply correlational. If causation is not established, or if it is not unidirectional, estimates of ‘effects’ from recursive models will be at the very least biased upwards, or may even be found to be spurious. If so, inferences regarding the substantive impact of voters’ for evaluations of competence and performance will be invalidated, and the empirical basis of the valence model will therefore be discredited.

For this purpose we estimate cross-lagged model parameters using the multiwave BES 2005–10 panel study to examine the direction of causality that is assumed in the model: that is, that responses to questions about the best-performing party and party leader influence party preference and voting intention, rather than party preference and voting intention influencing evaluations of the competence of parties and party leaders. The idea that voters’ party preferences are powerful influences on competence and performance assessments has been influential since at least the *American Voter* (Campbell et al. 1960; Green, Palmquist and Schickler 2002; Bartels 2002a). In this view, voters’ party preferences provide an interpretative lens through which politics is perceived, which biases their judgments so that their preferred party is seen in a positive light. We compare the relative magnitudes of these reciprocal effects. We also use this approach to examine heterogeneity amongst voters and test claims concerning the generality of the valence model, which rest on the idea that performance evaluations are easily employed heuristics that enable even less-informed and less-motivated voters to make political choices.

THE BES VALENCE MODEL

The BES valence model has three core elements: leadership assessments, party performance evaluations and party identification.\(^3\) The theory proposes that the key processes informing party choice are “comparative assessments of parties’ managerial capabilities and their

---

\(^2\) For other authors, valence politics has been a conditional outcome of the adoption of ideologically convergent strategies by parties (for example, Green 2007; Green and Hobolt 2008). Here we focus only on the assumptions and empirical implications of the Clarke et al. operationalization of valence.

\(^3\) We examine the appropriateness of including party identification in the model of valence voting elsewhere.
potential performance when attempting to deal with these problems. Voters maximize their utilities by choosing the party that they think is best able to deliver policy success in areas that concern them most” (Sanders et al. 2011, 289). The theory is operationalized and tested within a survey context by measuring performance evaluations and estimating their association with vote intention or reported vote. Thus “the valence indices for Labour and the Conservatives combine party leader evaluations, assessments of the party best able to deal with the most important problem facing the country, and evaluations of the prospective policy competence of a party with regard to the economy, the NHS, crime and terrorism” (Sanders et al. 2011, 304). Party choice is then regressed on these indices in either simple cross-section logit models, with all variables measured contemporaneously or near contemporaneously. Alternatively, in their most recent (and arguably most sophisticated) analysis, Sanders et al. (2011) use vote intention instead of reported vote and measure all covariates at the same time as the response, and include the lag of the dependent variable in their models.5

In general, the BES valence model is found to have substantially higher levels of predictive power than models containing, for example, issue-proximity scales or demographics. As a result, earlier specifications of the valence model emphasized its predictive superiority to models containing only demographics, particularly class voting (Clarke et al. 2004), while more recent publications have also stressed its superiority to models containing issue proximities, as “…voters are always likely to give priority to valence issues over spatial issues” (Clarke et al. 2009, 51). In the most recent specifications of the model, it has been claimed to be predictively superior to, though partly conditioned by, issue proximity (Sanders et al. 2011).

Underlying the arguments of the valence thesis is the idea that “people typically are not particularly interested in politics and wish to minimize the cognitive burdens of selecting a party” (Sanders 2011, 290). Valence theory “merely requires a voter to judge the relative merits of a very limited number of prominent individuals (party leaders) and to assess the relative performance capabilities of the main alternative parties in an area that the voter prioritizes” (Sanders et al. 2011, 290–1). Valence attributes are thus cognitive heuristics that minimize the cost of acquiring detailed information about politics.6 This further adds to the appeal of the model as an explanation of voting behavior of the large body of voters who are not especially informed about or motivated by politics. Valence considerations allow politically unsophisticated voters to choose a party.

WHY PARTY PREFERENCE MAY EXPLAIN PERFORMANCE EVALUATION

Why might valence not be the major explanatory force that its advocates have purported? As noted above, the idea that party preference provides a basis for assessing competence

---

4 The performance evaluations are measured in the post-election survey along with reported vote, but some other variables, such as demographics and party identity, are measured in a pre-election survey (Clarke et al. 2004). See also Clarke et al. 2009, 2011.

5 Unfortunately, the use of a lagged dependent variable in a logit model does not account for temporal dependence in the same way it would in a dynamic linear regression model (see Beck, Katz and Tucker 1998).

6 The same argument has, of course, been applied in the spatial modeling literature. From Downs (1957) onwards, ideologies and positions in policy continuums have been conceptualized as low-information, rational shortcuts that allow voters to make political choices. It is not clear why evaluating a plethora of performance indicators should be less cognitively demanding than holding an ideological position.
and performance is well established: “The strength of relationship between party identification and the dimensions of partisan attitude suggests that responses to each element of national politics are deeply affected by the individual’s enduring partisan attachments” (Campbell et al. 1960, 128). Bartels (2002a) likewise argues that the interpretation of party identification as a “running tally” of past satisfaction with a given political party’s policies is untenable, as perceptions of past performance are themselves shaped by pre-existing partisan affiliation. Much ensuing research has established that partisanship can influence responsibility attributions so that favored parties are seen to be responsible for good outcomes but not for negative ones (Abramowitz, Lanoue and Ramesh 1988; Feldman 1982; Peffley 1984; Rudolph 2003; Marsh and Tilley 2010).

Following Gaines et al. (2007), Tilley and Hobolt (2011) outline the mechanisms through which partisanship can influence evaluations. Voters can refuse to update their beliefs in the face of new information by, for example, disregarding unfavorable information on performance if they support the incumbent. Bartels (2002a) shows, for example, that many “strong Democrats” thought that inflation had increased during Reagan’s presidency, while it had in fact declined precipitously. Voters can also reinterpret a given level of performance in a more positive light (selective evaluation) or attribute responsibility for the level of performance differently (selective attribution).7 Inter-election panel analyses have likewise indicated that political preferences shape citizens’ perceptions of national economic performance (Evans and Andersen 2006; Evans and Pickup 2010), while other recent research by Ladner and Wlezien (2007) finds that voters’ expectations about which parties are likely to win forthcoming elections influence their prospective economic assessments. That these biases are consequential beyond the voting booth has also been established by Gerber and Huber (2009), who demonstrate that actual economic behavior is consistent with patterns of partisan differences in economic evaluations. Research into ‘motivated reasoning’ (Redlawsk 2002; Taber, Lodge, and Glather 2001; Taber and Lodge 2006) likewise implies that information gathering is unlikely to be accurate, as voters seek to bring their observations in line with prior opinions. The resolution of inconsistency in the case of performance evaluation therefore works to preserve longer-term beliefs associated with party preference rather than serve as a neutral information-gathering process, as implied in the valence model.


As noted above, the BES valence model assumes that the association between performance measures and vote intention/reported vote, sometimes net of prior vote intention (as in Sanders et al. 2011), is causal. The literature on party effects on evaluations implies that the direction of effects is highly unlikely to be one way. We therefore need to estimate recursive models. One way of addressing such a need is to attempt to identify relative strengths of the direction of influence between performance evaluations and party choice using two-step procedures with panel data.8 That such identification is inevitably arbitrary with cross-sectional data is well known, since it is difficult to find an instrument that is related to political

---

7 Tilley and Hobolt (2011, 328) conclude that: “voters do adjust their views of who is responsible in line with their partisanship. Although government partisans may acknowledge that the economy is in crisis, they also appear ready to absolve the government of responsibility for that crisis”.

8 For example, Markus 1982.
attitudes at time $t$ but is uncorrelated with the unobserved determinants of party preference at the same time (see Schickler and Green 1993). There are also survey context effects that militate against attempting to estimate the strength and direction of effects within surveys.

An extensive literature addresses how the questionnaire context is a source of bias and error in survey research (Schuman and Presser 1981; Schaeffer and Presser 2003) and how question proximity affects respondents’ interpretation and retrieval of information, which is typically referred to as priming (Tourangeau and Rasinski 1988; Tourangeau, Rips and Rasinski 2000). Within the study of electoral behavior there are numerous examples of priming derived from question proximity.9 Zaller’s influential (1992) monograph presents robust evidence of question-order effects and question-wording effects. Consistent with much research in cognitive psychology, Zaller and Feldman (1992) have proposed that most people respond to survey questions on the basis of whatever ideas happen to be salient and accessible to them at the time of answering. Such accessibility is conditioned by the content of preceding questions in a survey.

It is problematic, therefore, that the strongest evidential basis for the valence model has typically involved analyzing repeated batteries of questions that ask respondents (1) which party they think is most able to deal with the most important issue, (2) how well the party (first Labour and then the Conservatives) is performing, or would hypothetically perform, in various issue areas, followed by (3) the standard party identity questions and (4) vote intention or reported vote.10 Figure 1 presents the valence and party questions in a representative wave of the 2005–10 BES internet panel.

It can be seen that no less than 17 performance evaluation questions, plus three questions about party identification, precede the question on voting intention. There are almost no intervening questions. A battery on ‘leader images’ (affect towards party leaders) follows shortly after the vote intention questions.

The BES questionnaires thus contain an almost continuous stream of party-laden priming questions. These are questions that explicitly ask for an evaluation of a named party (or ‘the government’—the Labour Party at the time). It is thus unsurprising (and potentially uninformative about the motives underlying vote choice) that there should be a close link between answers to questions about party performance that explicitly label the party concerned, party identification and vote intention. The question batteries are so heavily primed with party cues that it is hard to imagine how there could be anything other than a high correlation between these various indicators of political orientation.

Such survey design effects represent a potential source of systematic measurement error, which is likely to seriously bias coefficient estimates. The implications of this bias have been demonstrated previously in research into economic performance evaluations, which are a key element of valence politics. Positioning questions on political attitudes immediately prior to economic evaluation questions influences respondents’ assessments of the national economy and their personal financial situation: it makes them more consistent with their partisan orientations (Sears and Lau 1983). Wilcox and Wlezien (1996) find that both political responses and economic evaluations are subject to contamination when related questions are placed in close proximity to each other. Other scholars have found similar priming effects in areas such as foreign policy (Krosnick and Kinder 1990).

---

9 For example, Sears and Lau 1983; Lau, Sears, and Jessor 1990; Krosnick and Kinder 1990; Bartels 2002b.

10 Data limitations mean that sometimes the models are estimated using a reduced set of indicators (for example, Clarke et al. 2011).
Our concern, then, is that the placement of explicitly party labeled performance evaluations adjacent to party preference questions will artificially inflate the correlation between responses. This hypothesized effect builds on a body of evidence indicating that people prefer to give consistent responses (Abelson 1968; Schwarz and Sudman 1992). Respondents condition their answers on their responses to previous questions. This preference for consistency produces a marked upward bias in the association between answers to proximal questions in social surveys. To the extent that this holds true, asking political questions in a context in which many other questions have cued the expected response is likely to produce an artifactual consistency between respondents’ performance evaluations.

Fig. 1. Question order and proximity in the BES internet panel questionnaire

---

11 This example is taken from 2010 pre-campaign questionnaire. All waves are very similar in layout, with inclusion/removal of some issues to reflect their changing salience, such as the financial crisis which didn’t exist at the start of the panel.
evaluations, and between those and their party preferences. Similarly, partisan priming derived from the content of the question batteries, with a persistent emphasis on which party the respondent thinks is best, can be expected to have made partisanship salient. Notwithstanding the problem of directional indeterminacy implied by the party effects literature referred to above, the strength of association between performance evaluations and party preference cannot be estimated with any confidence from a within-survey, cross-sectional analysis.

Our solution to these two sources of bias in estimating valence effects—the indeterminacy of causal direction and the upward bias of priming and questionnaire proximity—is to use panel studies to simultaneously model the effects of past performance evaluations on current party preference and of past party preference on current performance evaluations, while controlling for prior party preference and performance evaluations, respectively. This approach builds on work that uses cross-lagged structural equation modeling estimation to examine the relative strength of partisanship and issue proximities (Evans and Andersen 2004), policy preferences (Milazzo, Adams and Green 2012; Highton and Kam 2011), economic perceptions (Evans and Andersen 2006) and core political values (Goren 2005; Evans and Neundorf 2013). This method provides a more demanding and, in our view, less biased test of the impact of performance evaluations on party preference. It allows us to compare the relative importance of these competing ways of understanding survey responses and to examine whether some sorts of performance evaluations are of greater significance to party preference than others.

HYPOTHESES

If the BES valence model is valid, we would expect performance evaluations to have a significant lagged, net impact on party preference, but for party preference not to have such a lagged effect on performance evaluations.

HYPOTHESIS 1: The dominant causal relationship is from performance evaluations to party preference, rather than vice versa.

If there are significant party effects, this would imply that the observed association between performance evaluations and party preference is at least in part attributable to reciprocal causal influence, and that the effect estimates in valence models, which assume unidirectional effects, are biased upwards. If partisan conditioning arguments are valid, we would expect party preference to have a significant lagged, net impact on performance evaluations.

HYPOTHESIS 2: Party preference has a significant net effect on performance evaluations.

The use of performance evaluations as simplifying heuristics for people who are “not particularly interested in politics and wish to minimize the cognitive burdens of selecting a party” (Sanders 2011, 290) can be tested quite simply by estimating cross-lagged models separately for voters with different levels of attention to politics. If valence attributes are cognitive heuristics that minimize the cost of acquiring detailed information about politics, we would expect to find that valence effects are present among less politically interested and motivated voters as well as amongst the more attentive.

HYPOTHESIS 3: The net effect of performance evaluations is not conditioned by voters’ levels of attention to politics.
It could conversely be argued that valence voting is not that simple—it requires a high degree of attentiveness to the media and the public debate in order to perceive salient problems (which vary over time), to link them to party competence at different time points, to perceive and evaluate differences among parties when their cabinets change, and to make an electoral choice anew in each election without relying on party identification as a standing predisposition (which is arguably the most cost-effective heuristic of all). If this is the case, we might expect to find that party preference effects are stronger (and valence effects weaker) among less politically attentive voters.

**HYPOTHESIS 4:** The relative strength of effect of party preference on performance evaluations and *vice versa* is conditioned by voters’ levels of attention to politics: performance evaluations have weaker effects on party preference among less politically attentive voters, who instead use the heuristic of party preference to infer performance.

**DATA AND MODEL**

We use data from the 2005–10 BES internet panel, which was designed to produce representative samples of the adult population living in private households in Britain.\(^{12}\) Three waves were carried out in the 2005 and 2010 election years: before the election campaign, during the campaign and soon after the general election. There were follow-up surveys in 2006, 2008 and 2009, with no survey in 2007. Due to the change in the Labour leadership in 2007, we focus on the last three years of the panel: 2008, 2009 and 2010 (pre-campaign wave).\(^{13}\) We also use indicators from the 2006 wave to construct the lagged latent measures for *Labour party performance* and *Labour preference*.

To correct for measurement error associated with survey items, we use multiple indicators to estimate the latent constructs of Labour performance evaluation and Labour party preference at each time point.\(^{14}\) Latent Labour performance is modeled with four valence indicators: perceptions of the Labour government’s handling of the economy, Labour’s handling of the NHS,\(^{15}\) the Labour leader affect and whether the Labour party is best able to handle the most important issue facing the country. Latent Labour preference is estimated with the Labour party feeling thermometer, the Conservative party feeling thermometer\(^{16}\) and Labour vote intention. All indicators are standardized. It is a flexible feature of cross-lagged structural equation models that the errors associated with each indicator are allowed to be correlated across all time points (see Goren 2005; Highton and Kam 2011; Milazzo et al. 2012).\(^{17}\)

---

\(^{12}\) Detailed information about the survey is available at http://www.bes2009-10.org/.

\(^{13}\) We avoid using the final wave, as post-election responses are likely to be contaminated by the election outcome winners and losers adjusting their performance evaluations according to how they had voted in the election (Anderson et al. 2004).

\(^{14}\) Using multiple indicators for each latent construct helps attenuate measurement error bias (see Ansolabehere, Rodden, and Snyder 2008).

\(^{15}\) Responses are measured on a 1–5 scale from ‘very badly’ to ‘very well’. Questions on taxation and education were not asked in 2008 and 2009. The question on the financial crisis was not asked until 2009, and the question on Afghanistan was not asked until 2010.

\(^{16}\) Party and leader thermometer variables are measured on a 0–10 scale.

\(^{17}\) Allowing for the disturbances associated with the same survey item being correlated across waves helps “accommodate systematic error in the indicators of the latent constructs” (Berrington, Smith and Sturgis 2006, 28).
The four-wave linear dynamic structural equation model\(^{18}\) is presented in Figure 2. We estimate the factor loadings \((\lambda_k)\) of each indicator\(^{19}\) \((x_j \text{ or } z_m)\) on each latent construct \((\xi_j \text{ - 'Labour reference' or } \delta_m \text{ - 'Labour performance', respectively})\) at each time point \(t\). The estimated stability coefficients \((\beta, \text{ for Labour preference and } \gamma, \text{ for Labour performance})\) measure the effect of each latent construct in the previous period \((t-1)\) on this same construct in the current period \(t\). The higher the value of these estimates, the more stable the latent construct is over time. The cross-lagged effect parameters, \(\mu, \text{ and } \sigma\), are the effects of Labour preference in the previous period \((t-1)\) on Labour performance in the current period \((t)\), and the reciprocal effects of Labour performance \((t-1)\) on Labour preference \((t)\), respectively, for respondent \(i\) in period \(t\):

\[
\begin{align*}
\text{Labour preference}_{it} &= \alpha_1 + \beta_1 \text{Labour preference}_{i(t-1)} \\
&\quad + \sigma_1 \text{Labour performance}_{i(t-1)} + \xi_i \\
\text{Labour performance}_{it} &= \alpha_2 + \gamma_1 \text{Labour performance}_{i(t-1)} \\
&\quad + \mu_1 \text{Labour preference}_{i(t-1)} + \theta_i
\end{align*}
\]

Importantly, the cross-lagged parameters \(\mu\) and \(\sigma\) represent the net effects of one construct on another, controlling for the values of each dependent variable in the previous period. Omitting the stability parameters \(\beta\) and \(\gamma\) would bias the estimates of the cross-lagged effects upwards. Latent factors estimated using the 2006 indicators can have effects on their values at the next time point, but not the cross-lagged effects on the other construct. They are used merely as controls.

Although cross-lagged dynamic structural models offer a considerable improvement on cross-sectional analyses by using the natural time ordering of panel data, they do not solve all potential endogeneity problems. For instance, they do not allow for non-recursive simultaneous effects of Labour preference and government approval within the same wave. To the extent that the simultaneity bias occurs due to priming or context effects arising from questionnaire design, this is not necessarily a weakness of cross-lagged models. However, if the two constructs have purely exogenous simultaneous effects on each other, such reverse causality cannot be disentangled. We can establish ‘Granger causality’ only, by comparing the strengths of their lagged effects on each other, net of their respective stability effects.

RESULTS

Before estimating the full model presented in Figure 2, we evaluate the fit of the four-wave stability-measurement models for Labour performance (Table 1) and Labour preference (Table 2). Observed indicators load very highly on the respective latent factors. Both models fit the data very well, confirming that the chosen observed indicators tap into the underlying latent constructs.\(^{20}\) Both models show a high degree of factor stability over time, suggesting that voters’ assessments of government performance, and their preference for Labour, hardly fluctuate from wave to wave. The lowest stability coefficient is for 2006–08 in

---

\(^{18}\) All structural equation models are estimated using Maximum Likelihood in Stata 12.

\(^{19}\) The factor loading of the first indicator is constrained to 1.

\(^{20}\) The root mean square error of approximation is below 0.05 in Table 1 and just over 0.05 in Table 2. The ratio of the chi-squared statistic to the degrees of freedom is 5 in Table 1 and 9 in Table 2. The Comparative Fit Index measures are above 0.90 in both models.
the performance model, as would be expected given the party leadership change and the onset of the global financial crisis in the intervening period.

Table 3 presents the cross-lagged and stability estimates for the structural-measurement model in Figure 2, with the factor loadings omitted for clarity. Both latent constructs were remarkably enduring between 2006 and 2008, 2008 and 2009, and 2009 and 2010, although the stability estimates were significantly larger for Labour preference than for performance assessments. This finding suggests that the feelings about the majority party were, on the whole, more stable over time than the evaluations of government performance. The estimated cross-lagged structural coefficients of 0.33 (p < 0.001) and 0.31 (p < 0.001) indicate that performance evaluations and party preference, respectively,
both had statistically significant positive effects on each other between 2008 and 2009, net of their stability effects. This is evidence of a reciprocal relationship between the two constructs during this period. Thus respondents who held more positive views of the incumbent government’s performance in 2008 reported a substantially stronger preference for Labour in 2009, on average, controlling for their prior support for Labour. These reciprocal effects are not statistically significantly different from each other, suggesting that between 2008 and 2009, neither performance evaluations nor party preference dominated. In contrast, the lagged effect of performance evaluations on party preference is no longer significant in the 2009–10 period, while party preference continues to have a significant positive lagged effect on performance evaluations with the estimated coefficient of 0.37 (p < 0.001). The difference between the two cross-lagged estimates is significant at
p < 0.05, suggesting a stronger party conditioning effect at the end of the electoral cycle.\textsuperscript{21}

Overall, the finding of significant net effects of party preference on performance evaluation in both periods (2008–09 and 2009–10), even after controlling for prior performance evaluations, gives support to Hypothesis 2 but not Hypothesis 1.

To check whether the performance model is disproportionately driven by leader assessments, we excluded the leader affect indicators in the measurement part (Table 4).

\textsuperscript{21} As a robustness check, we replace the NHS performance indicator with the analogous survey item about the government’s competence in handling immigration (although the immigration question has only been asked since 2008). The estimates are nearly identical to those reported in Table 3 (available from the authors on request).
The estimated effect of party preference in 2008 on government performance in 2009 is 0.37 ($p < 0.001$), compared with the reciprocal effect of lagged government performance on party preference of 0.20 ($p < 0.001$). The difference between these two coefficients is statistically significant at $p < 0.01$, suggesting that if leader evaluations are excluded from the model, the effect of party preference dominated in this period. In the main model in Table 3, in which leader evaluations were included as indicators of government performance, the reciprocal effects of party preference and government performance are not significantly different from each other. During the 2009–10 period, party preference continues to have a significant positive lagged effect on performance evaluations, while the reciprocal effect is still statistically indistinguishable from zero. Thus, Labour preference now has significantly larger cross-lagged effects on performance evaluations than the reciprocal effects of performance assessments on Labour preference in both 2008–09 and 2009–10. These findings indicate that leader evaluations strengthen the overall effect of the valence indicators on party preference.22

Thus it would appear that leader evaluations are less likely to be conditioned by party support than competence measures.23

To evaluate Hypothesis 3, we re-estimate the original model (including leader assessments) separately for respondents who reported higher attention to politics in the 2005 pre-campaign wave (7–10 on the 0–10 scale)24 and those who reported lower attention to politics. Table 5 shows the estimates for the sub-sample of the respondents with higher values of reported attention to politics. The cross-lagged (structural) coefficients are very similar to those based on the full sample: both performance assessments and Labour preference have significant positive reciprocal lagged effects of

---

**Table 5** Cross-lagged Labour Preference and Performance Model (High Attention to Politics)

<table>
<thead>
<tr>
<th></th>
<th>2006–08</th>
<th>2008–09</th>
<th>2009–10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stability coefficients</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labour preference</td>
<td>0.95</td>
<td>0.88</td>
<td>1.26</td>
</tr>
<tr>
<td>Labour performance</td>
<td>0.81</td>
<td>0.63</td>
<td>0.48</td>
</tr>
<tr>
<td><strong>Structural coefficients</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preference → Performance</td>
<td>0.34</td>
<td>0.40</td>
<td></td>
</tr>
<tr>
<td>Performance → Preference</td>
<td>0.24</td>
<td></td>
<td>$-0.31^{ns}$</td>
</tr>
<tr>
<td><strong>Model fit</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chi-squared (df = 304)</td>
<td>5,847</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFI</td>
<td>0.886</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.112</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>79,725</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1,463</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: maximum likelihood estimates, with robust standard errors. All parameter estimates significant at $p < 0.001$, unless indicated otherwise. Factor loadings, factor variances, error variances, error covariances, intercept estimates and disturbances omitted for clarity. All indicators are standardized.*

---

22 See also Evans and Chzhen 2013.
23 See also Evans and Andersen 2005; Johns 2010.
24 Using the measure of attention to politics from the 2005 wave ensures it is temporally prior to the party support and performance assessment indicators used in the 2006–10 model. Attention to politics exhibits a high degree of stability over time. The measure is skewed toward higher values, with 39 per cent of the respondents in the estimation sample reporting scores under 7 in 2005.
comparable size on each other in 2008–09. In 2009–10, only Labour preference has a significant lagged effect on the other construct. The effect of Labour preference is also significantly larger ($p < 0.01$) than the reciprocal effect of performance evaluations.

In the model estimated on the sub-sample of respondents with lower self-reported attention to politics (Table 6), none of the cross-lagged effects is statistically significantly different from zero, net of the stability effects. The stability coefficients themselves are large and significant, suggesting that respondents with lower attention to politics are less likely to update their performance assessments and party feelings from year to year, and rely on their previous political attitudes. This does not support Hypothesis 4, in which less politically attentive respondents were predicted to use party preference as a heuristic for responding to questions about performance. Neither, however, is it consistent with Sanders et al.’s (2011) argument that valence considerations provide an effective, simple heuristic for updating party preference among those who are not strongly motivated by (or interested in) politics (Hypothesis 3).

It should be noted that none of the estimated models so far includes party identification. The inclusion of Labour party identity as an observed exogenous factor in the structural model would lead to multicollinearity and, therefore, inflated standard errors for all structural estimates. Alternatively, including party identity in the measurement part of the model as an indicator for both latent performance assessments and latent party preference precludes model identification. Since the valence model (Sanders et al. 2011) uses party identity alongside government competence and leadership assessments as a predictor of vote, it could be entered as an indicator of latent performance evaluations in our model. However, party identification can also be conceptualized as a long-standing commitment to a party (for example, Campbell et al. 1960). Consistent with this latter view, we find that party identification taps effectively into the latent party preference, along with the party feeling thermometers and vote intention.

---

**Table 6**  
**Cross-lagged Labour Preference and Performance Model (Low Attention to Politics)**

<table>
<thead>
<tr>
<th></th>
<th>2006–08</th>
<th>2008–09</th>
<th>2009–10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stability coefficients</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labour preference</td>
<td>0.91</td>
<td>0.72</td>
<td>1.07</td>
</tr>
<tr>
<td>Labour performance</td>
<td>0.74</td>
<td>0.81</td>
<td>0.58</td>
</tr>
<tr>
<td><strong>Structural coefficients</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preference → Performance</td>
<td>$0.20_{ns}$</td>
<td>$0.30_{ns}$</td>
<td></td>
</tr>
<tr>
<td>Performance → Preference</td>
<td>$0.53_{ns}$</td>
<td>$-0.09_{ns}$</td>
<td></td>
</tr>
<tr>
<td><strong>Model fit</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chi-squared (df = 304)</td>
<td>3,313</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFI</td>
<td>0.873</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.103</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>50,865</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>927</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: maximum likelihood estimates, with robust standard errors. All parameter estimates significant at $p < 0.001$, unless indicated otherwise. Factor loadings, factor variances, error variances, error covariances, intercept estimates and disturbances omitted for clarity. All indicators are standardized.*

---

25 Milazzo et al. (2012) use the strength of Labour party identification as one of their indicators of latent ‘Labour attachment’, along with the Labour feeling thermometer.
The Labour preference measurement-stability model—identical to the model in Table 2 but including the strength of Labour party identification as the fourth indicator—fits the data well (see Table A1 in the Annex). Importantly, it fits better than the comparable Labour performance model with party identity as the fifth indicator (Table A2). Thus if we were to have included party identity in the analysis it would have been most appropriate to have included it as an indicator of Labour preference in the cross-lagged models. When operationalized in this way, Labour preference has a significant cross-lagged effect on performance assessment in both periods, but performance assessment has a reciprocal effect only in 2008 (see Table A3). This pattern replicates that observed in Table 3, in which party identity was not included as part of the party preference factor.

CONCLUSIONS

We have examined whether the direction of association postulated in Clarke et al.’s valence model of British voting from performance, competence and leader evaluations to party choice is empirically valid. Our findings indicate that in the earlier part of the electoral cycle, the observed association between competence measures and party preference is a consequence of their reciprocal influence, but the effect of latent party preference clearly dominates toward the end of the electoral cycle. Thus controlling for party preference in 2008, we find that performance evaluations in 2008 have a significant positive effect on party preference in 2009. Yet we also find a significant positive reciprocal effect of a similar size of past party preference on current performance evaluation in the same period. The lagged effect of performance evaluation in 2009 on party preference in 2010 is not significant, however, while past party preference continues to exert a substantial positive effect on current performance evaluation in 2010. As would be predicted from the literature reviewed in the introduction on the impact of party preference on performance evaluations, assessments of government performance are, to a significant degree, a product of party preference and cannot therefore be used to explain that preference. These estimates are shown to be robust to alternative specifications of observed indicators, such as omitting leader evaluations from the performance assessment model or including Labour party identification in the Labour preference model.

We additionally address the assumption in the BES valence model that government and leader evaluations are heuristics that voters use to minimize the cognitive burden of choosing a party. This implies that performance assessment effects should be observed among less attentive voters as well as the more attentive. We find, however, that respondents with lower self-reported attention to politics simply do not update their government competence assessments and party preference from one year to another. This suggests that, rather than acting as a cost-minimizing heuristic, competence and performance assessments require a higher degree of attentiveness to current affairs than is consistent with the general applicability of the valence model. The results for those who reported higher levels of attention to politics are in line with the estimates for the full sample, in that they are more likely to re-evaluate government and leader performance every year. However, as with the full sample, these assessments are strongly conditioned by prior party support.

26 The strength of Labour party identification is coded: 0 ‘not a Labour identifier’, 1 ‘not a very strong identifier’, 2 ‘a fairly strong identifier’ and 3 ‘a very strong identifier’.
These findings are inevitably subject to qualifications. First, we have only considered the 2005–10 British electoral cycle. This was a period of sharp decline in popular support for the party in government as well as a time of significant economic turbulence, which may limit the generalizability of the results. At the same time, analyses of this electoral cycle form the basis of claims concerning the predictive superiority of the valence model in the most recent publications on this topic (Clarke et al. 2011; Sanders et al. 2011). The BES valence model is believed to be no less valid in this electoral cycle than it is in other cycles. Our findings are therefore likely to be of general relevance with respect to the validity of the assumptions and procedures used to estimate valence effects.

Secondly, although the use of lagged models is an improvement on the strong assumptions of simultaneous, unidirectional causality embodied in cross-sectional analyses, they do not offer an irrefutable test of exogeneity of performance assessments and party preference. Cross-lagged models are becoming increasingly popular, but they come with an important limitation: only lagged effects can be estimated, with no allowance for simultaneous non-recursive effects. Thus, at most we can establish ‘Granger causality’ (Highton and Kam 2011); party preference is said to cause performance assessments at a point in time because there is a significant association between prior party preference and current performance assessments, controlling for prior performance assessments. Given the evidence of endemic context effects in surveys, however, this limitation is not necessarily a weakness. As discussed in the introduction, there is strong evidence that simultaneous effects in surveys are powerfully influenced by the measurement context.

Our view is then that the use of lagged structural equation modeling, with multiple survey items measuring the latent constructs of interest, has provided more realistic estimates of the effects of performance evaluations than has hitherto been the case. In contrast, the measurement approach that has been adopted by Clarke et al. does not support the inferences drawn regarding its superiority vis-à-vis other approaches to explaining party choice. Cross-sectional models seriously over-estimate the magnitude of the impact of performance evaluations on party preference. This derives in part from misinterpretation of the direction of effects between performance evaluations and party preference, but it is also clear that the demands of the BES valence model—with respect to voter updating and use of performance as a heuristic—are also incorrect. The degree to which this is a measurement problem or a theoretical problem cannot be ascertained with certainty. It is possible that ‘purer’, less contaminated measures of performance evaluations have stronger influences on party preference than vice versa. However, to the degree that the BES valence theory is empirically disconfirmed, it cannot be taken as “a powerful, possibly unassailable, explanatory account of the overall pattern of party choice” (Johnston 2011, 213).

It should be made clear that the re-analysis and critique presented in this article relates most precisely to the operationalization of valence advanced with considerable

---

27 Previous BES panel surveys do not include the government competence measures necessary to evaluate the valence model.

28 To make sure that our findings are not limited to the post-leadership change period, we also estimated a two-wave model using indicators from 2005 (pre-campaign wave) and 2006 only. The lagged effect of Labour preference is significantly larger than the reciprocal effect of performance evaluations, although both estimates are significantly different from zero (see Table A4 in the Annex). The fit of this model, however, is considerably worse than that of our main (2006–10) model.
influence within the discipline, particularly in Britain, by Clark et al. At the same time, we might want to consider what it suggests for empirical models of valence theories more generally. We would suggest, first, that if valence theory is to make progress, studies that implement its empirical implications need to take into account the effects of question proximity on estimates of valence effects. The best strategy for dealing with such context effects is to include unrelated questions, so called “buffer items”, between the questions for which associations between responses are being estimated. A further, particularly problematic feature of the operationalization of valence examined here is the extensive explicit cueing of parties in the content of questions. Voters may well be able to evaluate issue competence without using party as a heuristic, but unless questions are devised that remove partisan cueing, we cannot know if they do or not.  

Valence theories might also benefit from refining their assumptions about how voters evaluate competence. An innovative though as yet unpublished study (Wagner and Zeglovits forthcoming) using cognitive interviewing procedures seems to indicate that the demands of the valence model on voter sophistication are such that refocusing on generalized party competence evaluations as advocated by Green and Jennings (2012a), rather than issue-specific evaluations, might be a productive strategy. This would help to address, in particular, the failure of less attentive respondents to update their issue competence evaluations and link them with their party preference in ways the theory has assumed. Similarly, studies of issue salience (Johns 2010) and leadership evaluations (Evans and Andersen 2005) suggest that the latter is a more robust source of influence on party preferences than the former, a finding that is also echoed in our analysis. Leadership effects are more robust to the tests we have presented in this article than are party performance evaluations. A related study (Evans and Chzhen 2013) of defection from Labour during the 2005–10 electoral cycle likewise supports the argument that leadership effects are not spurious and do, at least under some conditions for some leaders, impact voter behavior. Rather than treating valence as a multiply defined construct, it could therefore be more effective in future research to examine the differential impact of its distinct components and discover which ones ‘do the work’.  

As a final, more general point, we should note that the inferential problems produced for survey researchers by the survey context are not confined to valence models (though these models do appear to be particularly vulnerable to concerns about partisan labeling). Questions of endogeneity and causal direction bedevil all models that measure subjective independent variables (for example, issue positions and issue proximity, ideological positions and proximity, values, partisan identity) in the same survey as dependent variables (for example, vote intention, reported vote). Raising the hurdle for assessing the validity of models of political choice more generally should, in time, put our explanations of voting on a stronger empirical footing.

---

29 In this respect, the ‘most important issue’ question has arguably greater face validity than explicitly party labeled competence questions, but studies indicate that even in this case their partisan endogeneity probably invalidates their use as competence evaluations (Wlezien 2005; Johns 2010).

30 ‘General competence evaluations’ refer to the use of salient issues on which the party has a strong reputation as heuristics for the provision of competence evaluations on less well-known issues. This addresses the problem of voters simply not possessing sufficient “information to evaluate government or potential party performance across all areas of policy” (Green and Jennings 2012a, 315–6).
REFERENCES


---

**ANNEX**

**Table A1**

**Labour Preference Stability Model (With Labour Party ID)**

<table>
<thead>
<tr>
<th>Factor loadings</th>
<th>2006</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labour thermometer</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Conservative thermometer</td>
<td>-0.63</td>
<td>-0.62</td>
<td>-0.60</td>
<td>-0.64</td>
</tr>
<tr>
<td>Labour vote intention</td>
<td>1.05</td>
<td>1.00</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>Strength of Labour party ID</td>
<td>1.07</td>
<td>1.03</td>
<td>1.02</td>
<td>1.06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.95</td>
<td>0.99</td>
<td>0.96</td>
</tr>
</tbody>
</table>

**Model fit**

<table>
<thead>
<tr>
<th>Chi-squared (df = 77)</th>
<th>570</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFI</td>
<td>0.989</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.052</td>
</tr>
<tr>
<td>BIC</td>
<td>68.094</td>
</tr>
<tr>
<td>N</td>
<td>2,408</td>
</tr>
</tbody>
</table>

**Note:** maximum likelihood estimates, with robust standard errors. All parameter estimates significant at $p < 0.001$, unless indicated otherwise. Factor variances, error variances, error covariances, intercept estimates and disturbances omitted for clarity. All indicators are standardized.
### TABLE A2  Labour Performance Assessments Stability Model (With Labour Party ID)

<table>
<thead>
<tr>
<th>Factor loadings</th>
<th>2006</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Handling the NHS</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Handling the economy</td>
<td>0.96</td>
<td>1.06</td>
<td>1.11</td>
<td>1.13</td>
</tr>
<tr>
<td>Labour best on most important issue</td>
<td>0.71</td>
<td>0.73</td>
<td>0.80</td>
<td>0.88</td>
</tr>
<tr>
<td>Labour leader thermometer</td>
<td>0.92</td>
<td>1.06</td>
<td>1.07</td>
<td>1.09</td>
</tr>
<tr>
<td>Strength of Labour party ID</td>
<td>0.79</td>
<td>0.80</td>
<td>0.77</td>
<td>0.83</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.90</td>
<td>0.98</td>
<td>0.96</td>
</tr>
</tbody>
</table>

#### Model fit

- Chi-squared (df = 137): 1,124
- CFI: 0.979
- RMSEA: 0.055
- BIC: 93,653
- N: 2,408

*Note: maximum likelihood estimates, with robust standard errors. All parameter estimates significant at p < 0.001, unless indicated otherwise. Factor variances, error variances, error covariances, intercept estimates and disturbances omitted for clarity. All indicators are standardized.*

### TABLE A3  Cross-lagged Labour Preference and Performance Model (With Labour Party ID as an Indicator of Labour Preference)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Labour preference</td>
<td>0.95</td>
<td>0.83</td>
<td>0.94</td>
</tr>
<tr>
<td>Labour performance</td>
<td>0.79</td>
<td>0.75</td>
<td>0.64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Structural coefficients</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Preference → Performance</td>
<td>0.23</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>Performance → Preference</td>
<td>0.29</td>
<td>0.07*ns</td>
<td></td>
</tr>
</tbody>
</table>

#### Model fit

- Chi-squared (df = 412): 10,116
- CFI: 0.889
- RMSEA: 0.099
- BIC: 146,234
- N: 2,408

*Note: maximum likelihood estimates, with robust standard errors. All parameter estimates significant at p < 0.001, unless indicated otherwise. Factor loadings, factor variances, error variances, error covariances, intercept estimates, and disturbances omitted for clarity. All indicators are standardized. Labour preference is estimated with: Labour party thermometer, Conservative party thermometer, Labour vote intention, and the strength of Labour party identification. Labour performance is estimated with: handling the NHS, handling the economy, Labour best on most important issue, Labour leader thermometer.*
### TABLE A4  Cross-lagged Labour Preference and Performance Model (2005–06)

<table>
<thead>
<tr>
<th></th>
<th>2005–06</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stability coefficients</strong></td>
<td></td>
</tr>
<tr>
<td>Labour preference</td>
<td>0.93</td>
</tr>
<tr>
<td>Labour performance</td>
<td>0.55</td>
</tr>
<tr>
<td><strong>Structural coefficients</strong></td>
<td></td>
</tr>
<tr>
<td>Preference → Performance</td>
<td>0.41</td>
</tr>
<tr>
<td>Performance → Preference</td>
<td>0.21</td>
</tr>
<tr>
<td><strong>Model fit</strong></td>
<td></td>
</tr>
<tr>
<td>Chi-squared (df = 67)</td>
<td>13,618</td>
</tr>
<tr>
<td>CFI</td>
<td>0.783</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.191</td>
</tr>
<tr>
<td>BIC</td>
<td>173,491</td>
</tr>
<tr>
<td>N</td>
<td>5,530</td>
</tr>
</tbody>
</table>

*Note: maximum likelihood estimates, with robust standard errors. All parameter estimates significant at p < 0.001, unless indicated otherwise. Factor loadings, factor variances, error variances, error covariances, intercept estimates and disturbances omitted for clarity. All indicators are standardized.*