Polit Behav (2010) 32:1–28
DOI 10.1007/s11109-009-9099-8

ORIGINAL PAPER

A Computational Model of the Citizen as Motivated Reasoner: Modeling the Dynamics of the 2000 Presidential Election

Sung-youn Kim · Charles S. Taber · Milton Lodge

Published online: 14 October 2009

Abstract A computational model of political attitudes and beliefs is developed that incorporates contemporary psychological theory with well-documented findings from electoral behavior. We compare this model, *John Q. Public* (JQP), to a Bayesian learning model via computer simulations of observed changes in candidate evaluations over the 2000 presidential campaign. In these simulations, JQP reproduces responsiveness, persistence, and polarization of political attitudes, while the Bayesian learning model has difficulty accounting for persistence and polarization. We conclude that "motivated reasoning"—the discounting of information that challenges priors along with the uncritical acceptance of attitude-consistent information—is the reason our model can better account for persistence and polarization in candidate evaluations.

Keywords Political attitudes · Motivated reasoning · Bayesian learning · Attitude updating · Computational modeling · ACT-R

How do citizens develop and change their political beliefs and attitudes? Two theoretical perspectives have dominated answers to this question. One view suggests that beliefs and attitudes are strongly influenced by socialization, develop inertia

Detailed simulation results are available from the first author upon request.

S. Kim
Seoul, South Korea
e-mail: syounkim@gmail.com

C. S. Taber (✉) · M. Lodge
Department of Political Science, Stony Brook University, Stony Brook, NY 11794-4392, USA
e-mail: charles.taber@stonybrook.edu

M. Lodge
e-mail: milton.lodge@stonybrook.edu

through time, and thus are not very responsive to contemporary information from the political environment (Campbell et al. 1960; Niemi and Jennings 1991). An alternative perspective postulates that political beliefs and attitudes continually change in light of current information and are highly responsive to contemporary changes in the political environment (Downs 1957; Page and Shapiro 1992). These two perspectives have been repeatedly mapped onto various controversies pitting persistence against persuasion (as in the debates over whether news matters or whether campaigns matter) or stability versus change (e.g., whether party identification is stable or not).

Scholars have proposed a number of formal models of individual behavior and examined their implications for the persistence and responsiveness of political attitudes. Most notably, Achen (1992) and Gerber and Green (1998) proposed Bayesian learning models of political attitudes in the context of party identification, where rational individuals update their evaluations of political parties via Bayesian assimilation of new information (for a recent elaboration, see Bullock 2009). These models, however, are limited in important ways. Most notably, as we will demonstrate, Bayesian models can account for the responsiveness of evaluations to new information but they cannot account for their persistence (or even polarization) in the face of contrary information. In a similar vein, Bartels (2002) recently found that partisanship drives fundamental biases in perceptions of a variety of political persons and events. Applying a Bayesian learning model to panel data, Bartels showed that Republicans' and Democrats' perceptions of even "factual" political data (e.g., changes in the inflation rate) are strongly influenced by their partisanship. A simple model of naïve updating of political preferences that does not account for such biases cannot account for the data. Here we provide a theoretical model of individual psychological dynamics that explains motivated reasoning in general as well as the specific role of partisanship in political information processing.

We develop a psychologically-informed theory of political attitude formation and change that integrates both on-line (Lodge et al. 1995) and memory-based information processing (Zaller 1992; Zaller and Feldman 1992; Tourangeau et al. 2000). We then represent this theory as a computational model named *John Q. Public* (JQP) [1] and compare its formal implications for the dynamics of candidate evaluation against Gerber and Green's (1998) Bayesian learning model. Specifically, we apply JQP and this Bayesian model via separate agent-based simulations to the empirically observed changes in candidate evaluations over the course of the 2000 presidential election as captured by the National Annenberg Election Survey (NAES; Romer et al. 2003).

It is worth noting at the outset that our purpose, and therefore our approach, differs from the norms of formal theory in political science. We do not seek the

[1] Our model is developed within the ACT-R cognitive architecture (*The Adaptive Character of Thought – Rational*; Anderson et al. 2004), which is a leading theoretical and modeling framework used in cognitive science for a wide range of learned behaviors, among them language comprehension, the recognition and recall of information, inferencing, the formation of beliefs, and the learning of complex skills. However, while ACT-R provides comprehensive, integrated sets of cognitive mechanisms for learning, it lacks affective mechanisms, which are essential to current theories of political information-processing, and it lacks specific mechanisms for preference updating. Consequently, much of our work developing JQP was devoted to building affective and updating mechanisms and integrating them with the cognitive processes in ACT-R.

simplest model that might account for the data we simulate in this paper. Rather our purpose is to build a model that can account for these data *while also accommodating what is known from social and cognitive psychology about memory and information processes*. We are essentially adopting the formal norms of cognitive science, where one seeks models capable of explaining a wide range of complex human behaviors in psychologically plausible ways. That is, models are expected to account for both new and already known phenomena.

A Theory and Model of Political Information Processing

Our theory integrates cognitive and affective structures and mechanisms into one framework: (1) an associative network representation of knowledge and attitudes in long-term memory (LTM), (2) activation and decay mechanisms for concepts in LTM, which determine what information is accessible for retrieval into conscious working memory (WM), (3) processes for the construction of attitudes from accessible information in memory, and (4) processes for the updating and expression of cognitive associations and attitudes. In this section we develop each of these sets of mechanisms, first axiomatically and then procedurally.

Knowledge and Attitude Representation in Memory

The foundation for our theory of memory processes is the classic cognitive learning paradigm most closely associated with John R. Anderson's architecture of cognition (Anderson 1983; Anderson et al. 2004). This theoretical system rests on four basic axioms:

Axiom 1, Modularity: The human cognitive system consists of relatively independent subsystems (modules) such as a central processing system, goal system, and memory system.

Axiom 2, Adaptivity and Efficiency: The human cognitive mechanism is adaptive to the structure of the external environment and has evolved to be an efficient, though not necessarily parsimonious, information-processing mechanism.

Axiom 3, Parallel and Serial Processing: Cognitive processes are a mixture of parallel and serial processing. Parallel processing operates rapidly and efficiently because multiple processes operate simultaneously, while serial processing is slower and less efficient because only one process may occur at a time (as is characteristic of conscious deliberation).

Axiom 4, Semantic Structure of Memory: Human long-term memory is semantically structured in associative networks.

Axioms 1 through 3 are embodied in the design of the ACT-R modeling framework that we used to build JQP and will not be discussed here (see Anderson et al. 2004). Memory structures and processes (Axiom 4), however, require further elaboration because our approach differs from the classic cognitive paradigm built into ACT-R. In particular, JQP brings evaluative affect center stage; one's likes and dislikes for "objects" in memory (e.g., leaders, groups, and issues) play a central role in our theory:

Axiom 5, Hot Cognition: Virtually all social concepts in memory are affectively charged. With repeated co-activation socio-political concepts become positively or negatively charged and this affective evaluation— positive, negative, or both, strong or weak—is linked directly to its conceptual representation in long-term memory (Abelson 1963; Lodge and Taber 2005).

By election eve, citizens who have been exposed to the campaign will have formed impressions of the candidates, parties, and issues, and these evaluations will come into play automatically on exposure to new information (Lodge and Taber 2000, 2005).

Figure 1 illustrates our theoretical framework for memory (Axioms 4 and 5), using part of the knowledge structure about George W. Bush of a typical, liberal

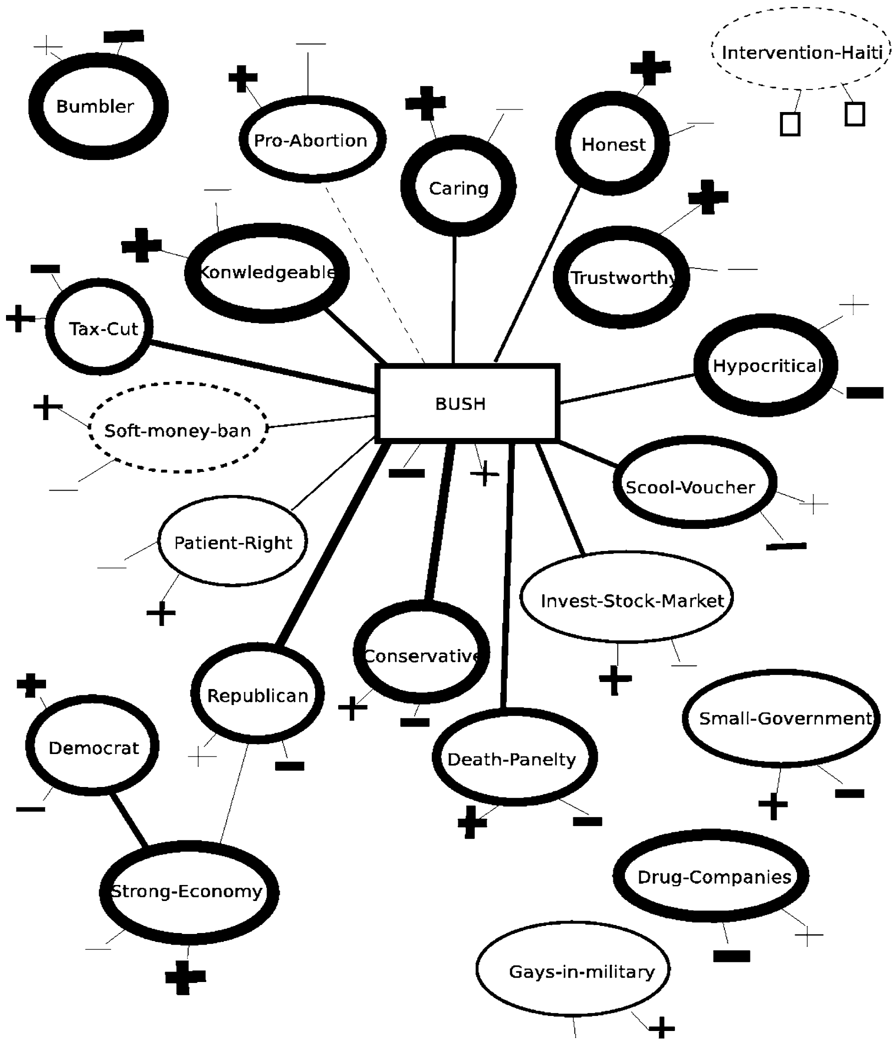

Fig. 1 Initial knowledge structure about bush for liberals

survey respondent in the NAES 2000 before the onset of the campaign. Each node or concept in memory is represented by an oval, the border-thickness of which varies to indicate differences in accessibility. For the conflicted liberal shown in Fig. 1, the traits “caring”, “honest”, “bumbler”, and “hypocritical” are all quite accessible, while the issues “patients’ rights” and “gays in the military” are less accessible. Associations between pairs of nodes are represented by connecting lines of varying thickness, which indicate their strength of association. So “conservative” and “Republican” are more closely associated with Bush in this respondent’s belief system than are Bush’s character traits. Plus and minus signs linked to the nodes represent positive and negative attitudes about the memory objects. A summary evaluation of an object may be obtained by combining the positive and negative valences (as when a survey respondent is asked for a thermometer rating of a candidate), but the theory can also represent ambivalence in cases where a node (e.g., “small government”) carries both positive and negative affect. Finally, every aspect of the initial knowledge structure—the particular object nodes and associations, the strengths of these nodes and associations, and the valences and strengths of evaluative tags attached to the nodes—change as citizens and agents respond to information throughout the campaign.

Accessibility of Memory Objects

Objects in long term memory (LTM) vary in their accessibilities (how easily and quickly they may be retrieved into conscious working memory [WM]) as a function of (1) the frequency and recency of past retrievals (practice and order effects), (2) the momentary activation received because of current processing of the node concept (as when reading the word “Bush” activates the “Bush” concept node), (3) activation spread to the node from associated concepts currently being processed (as when reading “Bush” activates an associated concept, “Republican”), (4) the degree of affective congruency between the node and information currently being processed (as when thinking about a negative concept like “terrorism” activates other negative concepts), and (5) the decay of activation levels through time and disuse (forgetting effects). All of these effects occur spontaneously and automatically, outside of conscious awareness.

These influences on accessibility, with the exception of affective congruency, are part of the classic cognitive paradigm (Axiom 4), and are built into the modeling framework within which we develop JQP. Affective congruency requires two new axioms and the development of additional procedures for the model.

Axiom 6, Primacy of Affect: Affect enters the processing stream before other thoughts and appraisals (Zajonc 1984), thereupon influencing the retrieval and interpretation of subsequent information (Taber and Lodge 2006).

Axiom 7, Affective Congruency: Information in memory that is affectively congruent with the information currently being processed is more accessible, while affectively incongruent concepts in memory are less accessible (Fazio 2001; Lodge and Taber 2005).

Hundreds of experiments in social and cognitive psychology document that affect enters the decision stream before cognitive considerations (Zajonc 2000). Neurological studies show that the “affect system” follows “quick and dirty” pathways that rapidly prepare us for approach-avoidance responses (LeDoux 1996, 2003). As we will show, the incorporation of affective mechanisms enables our model to represent motivated reasoning and account for both persistence and responsiveness in political information processing.

JQP’s activation mechanism, which determines the accessibility of a concept from LTM at a given moment in time, is an operationalization of the classical mechanisms of human recall (axiom 4), modified by the affective congruency effect on accessibility (axiom 6). Current activation of a given concept node is a function of its past activation, influences from associated concepts that are currently activated, and external stimuli that may trigger this concept.

More precisely,

$$A_i = B_i + \sum_{j=1}^{n} W_j\left[\left(S_{ji} - \ln\left(F_j\right)\right) + \gamma C_{ji}\right] + \sigma M_i + N\left(0, \frac{\pi s}{\sqrt{3}}\right) \qquad (1)$$

where A_i is the activation level of node i, B_i is the base level activation of node i (given in Eq. 2), W_j is the attention weight for node j (for $j = 1$ to n, where n is the number of nodes currently being processed), S_{ji} is the strength of association between nodes j and i, F_j is the number of nodes linked to node j, γ is a parameter governing the weight given to affective congruence, C_{ji} is a trichotomous indicator of affective congruency between nodes j and i, σ is a parameter governing the weight given to external activation, M_i is a binary variable indicating whether node i matches external information being processed, and N (0, $\pi s/\sqrt{3}$) is normally distributed noise with a mean of 0 and standard deviation determined by the parameter s (see Table 1 for summary of notation).[2]

Think of Eq. 1 in terms of its three primary subcomponents: (1) B_i includes the residual effects on the accessibility of node i of past processing and memory decay; (2) the complicated term following the summation sign represents the cumulative effects of other memory nodes on the accessibility of node i, broken into spreading activation, S_{ji}–ln(F_j), and affective congruency, γC_{ji}, [3] effects; and (3) σM_i captures the influence of external information on concept activation (e.g., reading “Bush” activates the “Bush” node). Note that both spreading activation and affective congruency effects are limited by the amount of focus (W_j) that may be given to node j, which is normally set at $1/n$ to represent the diminishing influence of any given concept when the number of concepts currently being processed increases.

[2] This notation for random noise is conventional in computational modeling because it makes transparent the manipulation of the normal density function.

[3] S_{ji} represents the strength of association from node j to node i. It is an increasing function of the number of times node j has sent activation to node i, and it is not symmetrical (Anderson 1993). $C_{ji} = 1$ when nodes j and i share the same valence (positive or negative), $C_{ji} = -1$ when they have different valences, and $C_{ji} = 0$ when either of them is neutral.

Table 1 Key JQP variables and parameters

Variables		Values
A_i	Activation of node i	Given in Eq. 1
B_i	Residual activation of node i	Given in Eq. 2
CA_i	Constructed evaluation of node i	Given in Eqs. 3a and 3b
OL_i	Evaluative tag attached to node i	Given in Eq. 4
W_i	Attention weight for node i	$1/n$, where n is number of objects in WM
S_{ji}	Strength of association from j to i	f (Number of times j has sent activation to i)
F_i	"Fan" dilution of activation from i	Number of nodes linked to i
C_{ji}	Indicator of affective congruency	1 if OL_i and OL_j same valence
		-1 if OL_i and OL_j different valence
		0 if either OL_i and OL_j are neutral (0)
M_i	Indicator of external match to i	1 if external information matches i
		0 if no match to i
T_{ij}	Processing time for activation decay	Time since i was processed jth time

Parameters		Value
γ	Affective congruence	2, estimated on NAES data
σ	External stimuli	5, after sensitivity analyses
s	Random noise in activation	0.1, from cognitive research
d	Memory decay	0.5, from cognitive research
δ	On-line versus memory-based processing	0.56, estimated on NAES data
ρ	Recency versus primacy in updating	0.94, estimated on NAES data

A second cognitive limitation built into Eq. 1 is the fan effect (F_j), which restricts the amount of activation that can be spread from node j to i when j is linked to a large number of other nodes. Equation 1 determines the accessibilities of objects in LTM and also forms the basis of the distinction between LTM and WM. Based on well-documented constraints on conscious processing (Anderson et al. 2004; Miller 1957), WM is modeled as a set of five limited-capacity buffers into which the most accessible objects in LTM are written.[4]

Though the accessibility of information in LTM decays rapidly, processing leaves residual activation, such that the more frequently and recently nodes have been activated in the past, the stronger will be their baseline accessibilities.

$$B_{it} = \ln\left(\sum_{j=1}^{m} T_{ij}^{-d}\right) \tag{2}$$

where B_{it} is the baseline activation of node i at time t, m is the number of times node i has been processed in the past, T_{ij} is the elapsed time since node i was processed the jth time, and d is a parameter representing the rate of memory decay that will

[4] The number of buffers could be manipulated to represent greater or lesser cognitive limitations, but we do not explore this implication here.

occur as time passes. So B_{it} increases with the number of times node i has been processed and with the recency of those activations, and it decays over time.

Memory-Based Attitude Construction

A basic task of democratic citizenship is the construction of attitudes about issues, parties, and candidates, as these underlie such consequential political behavior as voting and other forms of active participation (Kinder 1998; Taber 2003). But we know relatively little about how individual citizens actually form and revise their political beliefs and attitudes. We propose:

Axiom 8, Attitude Construction: Summary evaluations of objects (attitudes) are constructed by integrating the evaluations of objects that are accessible at the time of attitude construction (Zaller and Feldman 1992; Tourangeau et al. 2000).

Given the primacy of affect, it is clear that an evaluation of an attitude object will be first influenced by the evaluative tag attached to the object. This axiom posits that the evaluation is likely to be further influenced by any considerations about the object that come momentarily to mind. In JQP, this attitude construction process is implemented as a weighted average of accessible evaluative links.

$$\mathrm{CA}_i = (1-\delta)\mathrm{OL}_i + \delta\left[\sum_{j=1}^{n} a_j \mathrm{OL}_j\right], \text{for } j \neq i \tag{3a}$$

$$a_j = \frac{(A_j/A_i)}{\sum_j (A_j/A_i)}, \text{ for } j \neq i \text{ and } A_j > 0 \tag{3b}$$

where CA_i is the constructed evaluation of attitude object i, δ is a parameter that controls the influence of other currently accessible considerations (the js) relative to the evaluative tag already stored for object i (OL_i), OL_j is the existing evaluative tag for node j, A_i and A_j are the accessibilities of nodes i and j, a_j is the normalized accessibility of j relative to i, and n is the number of other accessible considerations at the moment of attitude construction.

The set of n considerations that enter this attitude construction process includes all objects associated in LTM with node i and all objects held in WM at the time of construction. The influence of each consideration is weighted by its relative accessibility (a_i), however, so effectively this set only includes objects with positive current activation ($A_j > 0$). Consequently, which considerations enter this construction process change dynamically through time. If no considerations are retrieved from memory at the time of attitude construction (i.e., $n = 0$), because either none is accessible at the moment or the attitude object has no associations in memory, then, $\mathrm{CA}_i = \mathrm{OL}_i$. In other words, when cognitions are unavailable or inaccessible, constructed attitudes reduce to on-line evaluative tags. Note that the parameter δ allows us to manipulate the degree to which our model represents on-line and/or memory-based models of information processing: $\delta = 0$ yields a purely OL model; $\delta = 1$ yields a purely MB model; $0 < \delta < 1$ yields hybrid models.

On-Line Processing of Attitudes

There is substantial evidence that attitudes and cognitive associations are routinely updated on-line, at the time that relevant information is encountered, with the evaluative implications automatically stored back to memory as OL tallies (Hastie and Pennington 1989; Lodge et al. 1995).

Axiom 9, On-Line Processing: Evaluations linked to objects in memory are updated continually and automatically upon exposure to new information, reflecting the influence of momentarily accessible information in WM.

Though the on-line (OL) model is well-established in social psychology, the failure to specify and test a specific updating rule has been rightly criticized (Kinder 1998). We offer such a rule in Eq. 4, which implements a form of anchoring and adjustment. In the model, attitudes constructed by Eqs. 3a and 3b are automatically integrated back into the evaluative tag associated with the attitude object.

$$\mathrm{OL}_{ir} = \sum_{k=1}^{r} \rho^k \mathrm{CA}_{jk}, \ \text{for } j \neq i \tag{4}$$

where OL_{ir} is the evaluative tag for node i that exists after processing the rth piece of information, ρ is a parameter that governs the weight of new relative to old information, and CA_{jk} is the attitude toward object j (as constructed by Eqs. 3a and 3b), which is the new piece of information associated with node i at processing stage k. Note that $\rho < 1$ implies the evaluative tag for node i becomes more persistent as more information about the object is processed.

Whenever new information is presented as input to JQP, a round of computation through these equations is performed. That is, new external information will cause JQP to compute a level of activation (Eq. 1) including a new baseline activation (Eq. 2) for every node in LTM, then retrieve the corresponding nodes into WM, then construct an evaluation of the new information that was presented (Eqs. 3a and 3b), and finally update the affective tag for the node associated with this new information (Eq. 4). After processing the information, the nodes held in WM are cleared. Processing stops when there is no new external information.

This set of cognitive/affective mechanisms is consistent with the distinction between automatic and deliberative processes. That is, they initialize and operate automatically. For example, when JQP reads and processes a sentence "Bush supports school-vouchers," much of the underlying processing, such as the spreading of activation to associated nodes and affective congruence effects, occur spontaneously outside of conscious awareness. The model becomes "consciously" aware of only the most accessible concepts and their associated affect when they are retrieved into working memory. JQP, like people everywhere, is only "aware" of the outputs of the process, not the process itself.

These mechanisms capture motivated skepticism for consistent initial beliefs when $\gamma > 0$, $0 < \delta < 1$, and $0 < \rho < 1$. That is, a set of coherent beliefs in JQP will tend to persist rather than fluctuate when affective congruency influences the spread of activation and memory retrieval (i.e., when $\gamma > 0$ in Eq. 1), when the model is neither purely OL nor MB (i.e., $0 < \delta < 1$ in Eqs. 3a and 3b), and when OL updating weights old information over new (i.e., $0 < \rho < 1$ in Eq. 4). Affective congruency is the key mechanism that drives the retrieval of considerations consistent with prior attitudes; attitude construction that mixes OL and MB processing promotes motivated reasoning because it is responsive both to the prior itself and to the biased set of retrieved considerations that can strengthen or even polarize that attitude; and attitude updating that favors priors will obviously promote persistence. JQP agents initialized with coherent priors (in this simulation these will be partisan agents) and appropriate parameter values will be motivated reasoners. Incoherent priors, on the other hand, would not drive the biased retrieval of considerations necessary for motivated reasoning.

An Illustration: How JQP Processes Survey Questions and Campaign Statements

Campaign information and survey questions are presented to JQP as English sentences, which the model parses word by word, retrieving corresponding concepts from memory, and updating knowledge and attitudes. JQP can represent any combination of political beliefs and feelings. For this illustration, JQP will be a liberal agent with the initial beliefs and attitudes shown in Fig. 1.

Consider first what happens when JQP is asked to answer the survey question, "How do you feel about George W. Bush?" Note that it is the JQP agents' average responses to this question and a similar one about Al Gore that are plotted over time in Fig. 2.

- First, the phrase "How do you feel about" triggers a routine that will generate an evaluation of the object that follows, in this case Bush.
- On reading "George W. Bush", the node for Bush receives activation computed by Eq. 1. Since the Bush node matches the input "Bush", M_{Bush} resets from 0 to 1, causing A_{Bush} to increase. With sufficient weight given to external information (governed by the parameter σ), the node Bush enters WM along with its directly associated affect.
- With Bush held in WM, other concepts associated with Bush (e.g., conservative, Republican, death penalty) now become more accessible because of the spread of activation from Bush (computed by a round of Eq. 1 for all nodes), with the increase dependent on the strength of association and degree of affective congruence between Bush and these nodes. So in addition to the cognitively associated concepts listed above, such affectively consonant nodes as hypocritical and bumbler also become more accessible, while evaluatively incongruent nodes—honest, pro-abortion, strong economy—become less accessible. Note that the influence of affective congruence is independent of semantic association. Some of these concepts, especially those that are both strongly associated and affectively congruent (e.g., conservative) will receive enough activation to enter WM.
- Using Eqs. 3a and 3b, JQP now constructs an attitude toward Bush that integrates the evaluative tags attached to Bush and other currently accessible

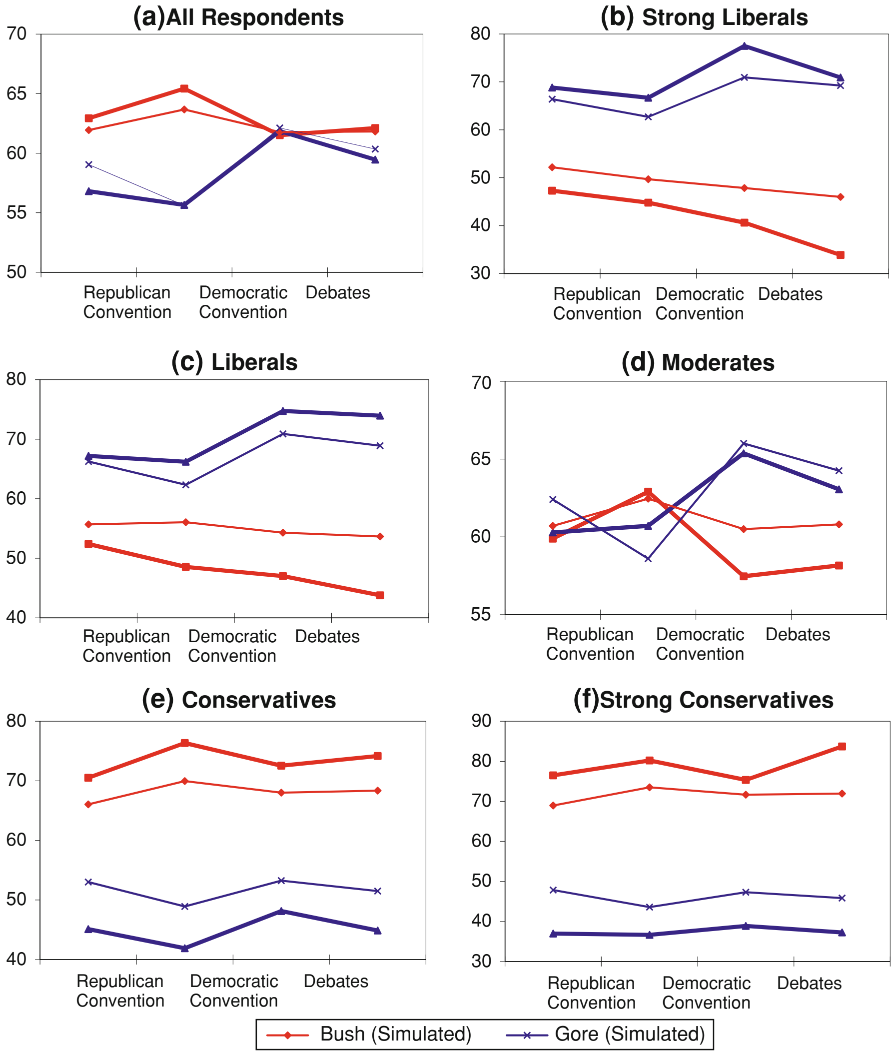

Fig. 2 The dynamics of candidate evaluation—survey versus simulation

information. Given the just-discussed spread of activation from Bush, such concepts as conservative, Republican, and hypocritical would significantly influence the attitude that is constructed, while others such as knowledgeable, caring, and strong-economy would have little influence. JQP would report this highly negative attitude in response to the question.

It is worth noting that we treated the survey question in this example as though it appeared in a vacuum. Since survey questions are generally embedded in an instrument with multiple items and opportunities for other influences (e.g., previous questions or race and gender of interviewer), there would normally be a history of processing that would affect how a given question is answered by JQP. Such a history alters the associations and on-line attitudes stored in JQP's long term memory. In addition, the act of processing a survey question like the one in this example would have some (negative) impact on the evaluative affect stored with the Bush node, since the newly constructed attitude will be integrated back into the on-line tally through Eq. 4.

Now consider a somewhat different example that illustrates motivated reasoning in JQP. Here, the campaign statement, "Bush is honest," is presented to JQP.

- Reading "Bush" increases the activation level of the Bush node and deposits it in WM as described above. Associated concepts and affectively congruent concepts receive activation from Bush as described above, and some of these concepts enter WM.
- After this initial recognition and processing of the word "Bush", JQP recognizes that the word "is" signifies that what follows will be about Bush. Any evaluative implication of "honest" will become associated with Bush.
- As JQP reads "honest", the node honest will receive activation and be deposited in WM along with Bush. Upon retrieval, the concept honest will influence the accessibilities of other objects in LTM in the same way that Bush did.
- JQP now constructs an evaluation of honest, but in the context of Bush. That is, if JQP were asked "How do you feel about honest?" in a vacuum, the response would likely be univalently positive. But an evaluation of honest computed by Eqs. 3a and 3b in the context of an activated Bush (and the mostly negative concepts associated with Bush in our example) will be very different indeed. The evaluation constructed in this context is the *subjective* implication of the campaign information "Bush is honest." For the liberal JQP in Fig. 1, this evaluative implication would certainly be less positive and perhaps even somewhat negative.
- The routine in JQP that was triggered by the word "is" updates the online tag attached to Bush using Eq. 4 and the evaluation constructed in the previous step for honest in the context of Bush.

In short, when the liberal JQP of our example reads the evaluatively inconsistent statement "Bush is honest," negative considerations such as "Bush is a conservative" and "Bush is a Republican" conspire with the preexisting negative attitude toward Bush to weaken (or potentially reverse) the positive implication of the information. By contrast, an evaluatively consistent statement like "Bush supports school vouchers" would feed into the already negative evaluation of Bush. Given strong enough negative priors for Bush and a dense network of consistently negative associations to Bush, our liberal JQP would likely polarize in the direction of her negative priors whether the presented information is objectively positive or negative. A conservative JQP would exhibit the opposite pattern of motivated bias.

In sum, when JQP processes a piece of campaign information, the structure of prior beliefs and attitudes influences how the information is subjectively perceived. And when asked to report an attitude about a candidate, information that favors priors will be more heavily weighted. In other words, JQP is a motivated reasoned:

it discounts discrepant information while accepting consistent information more or less at face value.

Simulating the Dynamics of Candidate Evaluation

Changes in the evaluation of a political candidate, group, or idea over time in JQP are determined by (1) the set of affective/cognitive *information-processing mechanisms* described in the last section, (2) initial *knowledge structures*, and (3) the flow of *campaign information* over time.[5] Individual voters differ in the second dimension, and they may differ in the third through selective exposure, so multiple JQP agents were employed to reflect these individual differences in the simulation of the 2000 election. Since it would be computationally impractical (and theoretically uninteresting) to completely specify the idiosyncrasies of each and every survey respondent, we developed the following straightforward simulation design:

- Based on responses to the NAES 2000 pre-election survey we identify initial knowledge structures for five self-identified ideological groups among the survey respondents—strong conservatives, conservatives, moderates, liberals, and strong liberals.
- We employ 100 JQP agents to represent a sample of the survey respondents. The initial knowledge structures of our agents are stochastically generated to replicate the distribution of ideological beliefs in the survey sample, yielding seven strong conservatives, 29 conservatives, 41 moderates, 19 liberals, and four strong liberals in the simulated sample. Though each agent falls within the range of beliefs associated with its ideological type, each also differs individually as a result of the stochastic generation process (explained below).
- We collected a measure of campaign information from the *New York Times*, the Long Island newspaper *Newsday*, and the *Wall Street Journal* relating to the major campaign events during the 2000 election. These events are the Republican and Democratic conventions and the three major candidate debates (treated as a single event because of their close temporal proximity). In order to test for potential media effects, we run separate simulations for the ostensibly liberal *New York Times* and *Newsday* versus the conservative *Wall Street Journal*.
- Finally, we conduct the simulation: Our stochastically generated sample of 100 agents answer the NAES 2000 survey questions before the campaign, then “read” distilled news accounts of the three campaign events, answering the

[5] To test the internal validity of the model (i.e., whether the model behaves as expected in our theory), we conducted a series of purely formal computational experiments (Kim 2005) in which the model successfully reproduced practice, recency, and spreading of activation effects on recall; cognitive and attitude priming effects; question order and wording effects in the survey response; both on-line and memory-based processing; and the ability to learn by adjusting beliefs and attitudes in response to campaign events. These tests were essential to establish that the model is in fact consistent with a wide range of “well-known phenomena” from the empirical cognitive literatures. These tests are available from the first author.

NAES survey again after each event. This procedure simulates a four wave panel across the course of the campaign that can be compared to actual NAES responses at these time points. Because it includes random components (the stochastic generation of knowledge structures and the normally distributed noise in Eq. 1), the simulation is repeated so that we in fact generate 100 simulated panels of 100 agents each.

Initial Knowledge Structures

Our theory defines knowledge structures (see Fig. 1) in terms of a set of concepts and associations among concepts, with evaluative tags linked to each concept, all of which vary in strength. Our primary source for the information used to generate initial knowledge structures was the early cross-sectional data in the 2000 NAES, collected during the primary season between December 1999 and July 2000, that is, before the GOP convention. For each of the five self-identified ideological groups, we obtained the distribution (mean and standard deviation) for attitudes toward the candidates, parties, and issues, as well as perceptions of the candidates' traits and issue positions.[6]

To generate the initial knowledge structures, each of the 100 agents was created and assigned to one of the five ideological groups (according to the frequencies reported earlier from the survey data). Memory objects (nodes in Fig. 1) were then created for every political concept (e.g., Bush, abortion, tax cut, etc.) and trait concept (e.g., honest, trustworthy, etc.) asked in the NAES 2000 survey. Initial baseline accessibilities of the political memory objects were set using the response rates for each survey item as a proxy for the frequency and recency of use. Semantic associations (links in Fig. 1) among memory objects were created when such associations were indicated in the NAES data for the given ideological subgroup. That is, the strengths of links among concepts were set according to the subgroup's mean response to such survey items as "Is Gore trustworthy?" or "Does Bush support a tax cut?" Initial evaluative affect for each candidate, group, and issue was assigned according to the distribution of feeling thermometer responses for the given ideological subgroup. In short, initial accessibilities, associations, and evaluative tags were generated stochastically following the distributional properties (mean and standard deviation) from the survey data for the given ideological subgroup. Though agents fall within the "normal" range of beliefs for their ideological subgroup, they differ in their specific beliefs (links and accessibilities) and attitudes (evaluative affects), so that each draw of 100 agents forms a unique sample.[7]

[6] For the trait perceptions to be useful, we need to assign a value to each trait. That is, when a respondent says "Bush is trustworthy," we need to have some idea how positive this attribution is. Fortunately, most human traits and other general concepts have been normed by large samples of respondents, so we can scale the evaluative implications of a given trait (positivity, negativity, intensity, etc.) by consulting the Affective Norms for English Words (ANEW; Bradley and Lang 1999) and N. Anderson (1968), which provide means and standard deviations for a large number of trait concepts.

[7] Recall also that transient random noise is added to the activation levels of all objects in memory each time an object is retrieved (Eq. 1), which also approximates individual differences in processing.

Campaign Information

Campaign information was obtained from *Newsday*, *New York Times*, and *Wall Street Journal* accounts of the Republican Convention (7/31-8/3), the Democratic Convention (8/14-8/17), and the three debates (10/3, 10/11, 10/17). These news accounts, we assume, roughly represent the information available to the NAES 2000 survey respondents. Since only about 10% of NAES respondents report reading "a great deal of" newspaper coverage of the campaign, our strategy was to select the single most extensive news story from the given paper for each event and code that story for the simulation. Our initial simulations will use the reputedly more liberal news sources (*Newsday* and *New York Times*), with a follow up simulation using a more conservative source (*Wall Street Journal*).

The information in these articles was coded into a format accessible to JQP: simple campaign statements attributable to some known actor (e.g., "Bush says Gore is dishonest"). Recognizing that many subtleties are smoothed away in the process, we developed a procedure to recover the gist meaning of each paragraph of a given news article. Each paragraph of each news story was scanned for simple assertions about the two major candidates, and that information was extracted as gist statements. For example, the statement (*Newsday*, third debate, 10/18/2000), "Gore pointed to Bush and said, 'If you want someone who will spend a lot of words describing a whole convoluted process and then end up supporting legislation that is supported by the big drug companies, this is your man,'" was coded "Gore said Bush supports Drug Companies." All qualifications were ignored, all modifiers excised, reducing the complex text to a bare skeleton.

This focus-on-the-gist procedure has several benefits for our purposes. First, it is conservative in the sense that it provides less information to JQP than was potentially available to real citizens (though perhaps as much as most citizens actually processed). Second, our approach minimizes the subjective interpretation process, which is the part of content coding most fraught with error. Finally, there is evidence that citizens do indeed process the gist meaning of campaign statements and ignore even not-so-subtle qualifications (Hamill and Lodge 1985; Lodge et al. 1995; Taber and Steenbergen 1995). Following this procedure, the first author reduced all article paragraphs to gist meaning, and the other two authors checked and on rare occasions modified these decisions.

The Agent-Based Simulation

The empirical dynamics to be explained in this paper are changes in feeling thermometer ratings of Bush and Gore for each of the five ideological groups over the course of the 2000 campaign, and in particular across four time points: before the campaign began, after the Republican convention, after the Democratic convention, and after the three candidate debates. We will compare the trajectories of candidate evaluations for our simulated agents to the trajectories for NAES respondents.

Because of the stochastic components in the model, we repeat the simulation 100 times. One may think of the 100 simulations as representing the sampling distribution and each simulation as representing an individual sample. Given the significant computational demands of this procedure, we ran the simulation on the Teragrid Supercomputer (http://www.teragrid.org). In addition to the candidate evaluations of Gore and Bush taken at four simulated time points in the campaign, on which we will focus, we also kept the complete trace of all internal psychological dynamics (such as momentary changes in internal attitudes on reading each new piece of information).

Parameter Values

Computational modelers treat parameters in several different ways: (1) parameters that are of particular interest to some hypothesis under examination may be manipulated experimentally; (2) parameters may be estimated based on model fit with empirical data; or (3) parameters may be manipulated systematically to determine the sensitivity of model behavior and to find reasonable ranges of values within which a model behaves as expected.

There are six free parameters in JQP. Two of these that are common to all ACT-R models were set to values that have previously been empirically estimated in the cognitive science literature (Anderson et al. 2004): s, which governs the amount of random noise in Eq. 1, was set to 0.1; and d, which regulates the rate of memory decay in Eq. 2, was set to 0.5. Three parameters unique to JQP were estimated to provide optimal fit to the NAES 2000 data on candidate evaluations: γ, which governs the influence of affective congruence in Eq. 1, was set to 2; δ, which controls the degree of on-line versus memory-based processing in Eqs. 3a and 3b, was set to 0.56; and ρ, which controls the relative weight of early versus recent information in Eq. 4, was set to 0.94.[8] The final parameter, σ, which governs the direct influence of external stimuli on activation in Eq. 1, was set to a value of 5 after sensitivity analyses. It is important to emphasize that model behavior in JQP is not very sensitive to reasonable variation around these parameter values (these sensitivity analyses are available from the first author). It should also be pointed out that the optimized values for γ, δ, and ρ will implement motivated reasoning for agents with minimally consistent prior beliefs (again, we did not choose these key parameter values, but rather they were estimated on the data).

Results for the JQP Simulations of the NAES 2000 Data

Table 2 reports several correlational measures of overall model fit to the NAES data, based on simulations using the *New York Times* and *Newsday* coverage of the campaign events. Section A in the table reports correlations over the course of the campaign events, averaged over the 100 simulations, for evaluations of Bush and

[8] The average of correlation between the actual and simulated evaluations over time and that across groups was used as a fit measure. However, the same qualitative results were obtained with wide ranges of parameter values.

Gore, broken down by ideological group. These correlations, averaging about 0.90 for all voters, represent how well the simulations track changes in candidate evaluations over time in the NAES sample. Section B reports correlations between simulated and actual evaluations across the ideological groups, broken down by time period. This measure, averaging 0.91, represents how tightly the simulation results fit the empirical distribution of candidate evaluations across the ideological spectrum. Finally, section C reports the distribution of results across the 100 simulations. As should be expected, there were several outlier simulations that did not track the actual results as well as others (the worst fit was 0.69 over time and 0.56 across groups), but the standard deviations are reasonably small so we can be confident that the 100 simulations clustered together rather tightly.

Having demonstrated a strong and robust model fit across time and across the ideological spectrum, we turn to a closer examination of dynamics for the simulation. Here, we examine a single simulation run to simplify the discussion. Figure 2 plots the evaluations for the 100 simulated respondents in this run against the actual NAES evaluations, with panel (a) comparing evaluations of Bush and Gore for all voters across the campaign, and the remaining panels breaking down

Table 2 JQP model fit—correlations between survey and simulated evaluations

Ideological group	Bush	Gore	Average
A. Average correlations across time			
Strong liberal	0.89	0.87	0.88
Liberal	0.90	0.89	0.90
Moderate	0.90	0.76	0.83
Conservative	0.96	0.93	0.95
Strong conservative	0.50	0.55	0.52
All voters	0.91	0.89	0.899
Time	**Bush**	**Gore**	**Average**
B. Average correlations across groups			
Initial	0.88	0.92	0.90
GOP convention	0.92	0.91	0.91
Dem. convention	0.92	0.92	0.92
Debates	0.92	0.91	0.92
Average	0.91	0.92	0.91
	Over time	**Over groups**	**Average**
C. Distribution of model fit over 100 simulations (all voters)			
Minimum	0.69	0.56	0.63
Maximum	0.96	0.99	0.92
Median	0.91	0.94	0.88
Mean	0.899	0.91	0.86
Standard deviation	0.04	0.08	0.05

the comparison by ideological group. We will compare the simulated and empirical trajectories in terms of starting positions, direction and magnitude of changes, polarization of evaluations, and ending positions.

First, across all ideological subsamples, JQP got the initial relative evaluations of Bush and Gore right. In no case did the simulated sample favor the wrong candidate after model initialization. Moreover, with the exception of the 41 moderate agents who on average liked both candidates more than their survey counterparts did at the start of the campaign, initial simulated evaluations were very close to actual evaluations. Looking at these data, it is clear that the stochastic initialization procedures produced a simulated sample that corresponds closely with the NAES sample before the start of the campaign.

The critical test for JQP is how well the comparison holds up as the evaluations of both agents and citizens change on exposure to campaign information. Setting aside the moderates for the moment, we find that all changes save one were in the right direction (liberal agents' evaluations of Bush after the GOP convention rose slightly, though the liberal NAES respondents said they liked Bush less after the GOP convention). That is, the sample of artificial agents overwhelmingly responded to campaign information by moving in the same directions as the NAES respondents, reproducing the same qualitative patterns in nearly all cases. And even the trajectories for moderates support the model, once we recognize that the changes after the GOP convention represent a correction to the poorly initialized moderate agents. In fact, after this first convention is processed, the simulated moderates converge quite tightly on their empirical counterparts.

In addition to being in the same direction, most changes in the simulated trajectories are of comparable magnitude to changes in the real-world trajectories. Moreover, agents' evaluations become more polarized (i.e., more extreme) after the campaign, though to a lesser degree than those of NAES respondents, in all ideological groups as did the real respondents (more on this to come). Finally, all simulated ideological groups approached Election Day with the preferences over candidates ordered correctly. In short, the broad qualitative picture to emerge from the simulation is identical to that from the NAES 2000 survey.

Actual survey respondents have potentially richer and more diverse sources of media information than did JQP, and they have some control over what they see and who they listen to. While JQP "read" the relatively liberal newspapers *Newsday* and the *New York Times*, real-world voters could read different newspapers, watch TV news, talk to their friends, and browse the internet. One might expect liberal and conservative citizens to read different newspapers and get a different slant from them, and these differences are at best only partially captured in the simulation reported above. Table 3 looks at how closely JQP reproduces the dynamics of observed candidate evaluations when agents read *Newsday* and the *New York Times* versus the *Wall Street Journal*.

Section A in Table 3 reports across-time correlations between simulated and empirical evaluations, averaged across the two candidates and over the 100 simulations, for each media source, broken down by ideological group. Clearly, the model tracked quite well the observed changes in evaluations over the course of the campaign for both media sources. Intriguing changes in model fit were also

Table 3 Comparison of liberal and conservative media (100 simulations)

	Newsday/New York Times		Wall Street Journal	
	Mean	Std. Dev.	Mean	Std. Dev.
A. Across time correlation—responsiveness				
Strong liberal	0.88	0.07	0.66	0.2
Liberal	0.90	0.1	0.72	0.12
Moderate	0.83	0.04	0.69	0.06
Conservative	0.94	0.03	0.86	0.03
Strong conservative	0.52	0.11	0.79	0.07
All voters	0.899	0.04	0.77	0.05

	NAES data	Newsday/NYT	Wall Street Journal
B. Magnitude of change—persistence			
Strong liberal	5.27	2.71	2.28
Liberal	4.0	2.63	2.27
Moderate	2.39	2.06	1.99
Conservative	2.51	2.0	2.01
Strong conservative	2.39	2.1	2.18
All voters	2.24	1.96	1.95

	NAES data		Newsday/NYT		Wall Street Journal	
	Before	After	Before	After	Before	After
C. Polarization in evaluations of bush and gore: candidate differential before GOP convention versus after the debates						
Strong liberal	−21.5	−37.0	−13.7	−20.1	−13.7	−17
Liberal	−14.8	−30.2	−10.4	−16.0	−11	−13.8
Moderate	−0.4	−4.9	−2.3	−4.4	−1.7	−1.22
Conservative	25.4	29.3	12.6	14.8	13.3	17.1
Strong conservative	39.5	46.5	20.1	24.0	20.6	25.7
Average (absolute)	20.3	29.6	11.8	15.9	12.1	15

observed, however. Note the pattern of correlations broken down by ideological groups: strong liberal and liberal agents track better when reading the liberal source, while strong conservative agents do appreciably better when reading the more conservative *Wall Street Journal*.

Section B in the table compares the amount of persistence in candidate evaluations across campaign events for real and simulated voters, as measured by the standard deviations of the distributions of evaluations across time, for each media source, broken down by ideological group. Small standard deviations indicate relative persistence in evaluations over time, while larger standard deviations show more change. Though the evaluations of real voters in the NAES survey changed somewhat more than those of simulated agents, especially for liberal respondents, the striking finding is the overall similarity in persistence (e.g., a standard deviation of 2.24 for all NAES respondents compared to 1.95 on a 100 point scale). In short, both real and simulated voters show substantial persistence in their evaluations over the course of the campaign.

Section C reports the actual and simulated differences between the evaluations of Bush and Gore (that is, Bush minus Gore), before the GOP convention and after the debates, for each media source, across ideological groups. This provides a measure of the degree to which the real and simulated samples either moderated or polarized over the campaign. As can be seen, the model reproduces the observed polarization of candidate evaluations. On average, the NAES respondents polarized from a Bush-Gore difference of 20.3 points (on a 100 point scale) at the outset to 29.6 by the end of the campaign, while the simulated agents polarized from 11.8 to 15.9 when using a liberal news source and from 12.1 to 15 using a conservative source. Once again, the differences were not as large for the simulation, but they were consistently in the same direction.

Overall, these simulation results fit the data very well, both qualitatively and quantitatively—JQP closely reproduces the observed responsiveness, persistence, and polarization of candidate evaluations of the survey respondents. Correlations between simulated and actual changes in candidate evaluations over time were consistently high, especially when selective exposure to an ideologically congenial media is taken into account. The changes in actual and simulated evaluations over time were of comparable magnitudes. And the agents' candidate evaluations became more polarized after the election campaign as did the real survey respondents. This was true for nearly all ideological groups, despite the fact that no individual-specific or ideological group-specific parameters were employed in the simulations; that is, *the mechanisms themselves and the parameter values were the same across all individual agents and all ideological groups*. The campaign messages processed by all agents were also identical (within a given media simulation). Only the initial knowledge structures for agents differed, and those were stochastically generated based on empirical data. In short, a single theoretical model simulated closely changes in evaluations of the two candidates over the course of a presidential campaign across five ideological groups. We think that this is a remarkable result that strongly supports the empirical validity of the John Q. Public model of political attitude formation and change in the context of a political campaign.

Moreover, these results are not very sensitive to reasonable variations in parameter values. Observe that as long as $\sigma > 0$, $\delta < 1$, and $\rho > 0$, campaign information matters; that is, JQP will respond to external information. Also, as long as $\delta > 0$ and $\gamma > 0$, prior beliefs and attitudes will bias the processing of information in the direction of priors. This implies broad ranges of parameter values within which the fundamental results of this simulation would obtain, though the *magnitude* of the effects would vary. In general as γ and δ increase, motivated biases become stronger and priors become more persistent. As σ and ρ decrease the responsiveness of the model to new external information weakens. The parameters s and d do not play a significant role in the results we report, though both are important to other cognitive functions not discussed in this paper.

A Comparison with a Bayesian Learning Model

At this juncture it would be useful to compare JQP with other learning models in assessing its empirical validity. Gerber and Green (1998) proposed a variant of a Bayesian updating model based on the Kalman filter algorithm, which encompasses Achen's learning model (1992) as a special case. Though Gerber and Green were particularly interested in how citizens update beliefs about the party differential, their model is general enough to be applied to candidate evaluations over an election campaign under suitable assumptions. This model is arguably the most sophisticated Bayesian learning model employed in this literature and so it is a reasonable choice for comparison with JQP, but it is important to keep in mind that our results using this model cannot be generalized to all Bayesian learning models.

Using similar notation to Gerber and Green (1998), suppose that there is some true underlying quality of a candidate (α_t) that forms the basis of citizen evaluations and that this quality changes through time according to the process:

$$\alpha_t = \gamma\alpha_{t-1} + \eta_t, \tag{5}$$

where $0 \leq \gamma \leq 1$ and $\eta \sim N(0,q)$. Changes in α occur because of both constant (γ) and random (η_t) factors.

Suppose that a citizen tries to estimate this underlying quality given a stream of information about the candidate (y_t).

$$y_t = \alpha_t + \varepsilon_t, \tag{6}$$

where $\varepsilon \sim N(0,h)$. That is, the citizen forms an estimate of α_t, which we denote $\hat{\alpha}_t$, updating her prior estimate $(\hat{\alpha}_{t-1})$ based on the current observation y_t. The Kalman filter algorithm for this quantity is based on the assumption that voters combine their prior beliefs with new information in an optimal fashion, by which is meant they minimize the expected squared error of their estimates.

$$\hat{\alpha}_t = \gamma\hat{\alpha}_{t-1} + K_t(y_t - \gamma\hat{\alpha}_{t-1}), \tag{7}$$

where $K_t = (\gamma^2 P_{t-1} + q)/(\gamma^2 P_{t-1} + q + h)$ and $P_t = hK_t$. Substantively, K_t is the weight given to a new observation (y_t) in obtaining the current estimate of the variable ($\hat{\alpha}_t$) based on the prior estimate $(\hat{\alpha}_{t-1})$ and the new observation, and P_t is the uncertainty associated with the estimate $\hat{\alpha}_t$ at time t, that is, the current variance of $\hat{\alpha}_t$. Assuming normally distributed errors, this algorithm gives the lowest mean squared error and thus provides an "optimal" rule for updating estimates of candidate qualities which will drive evaluations. This updating approach can be derived from Bayes' rule, and so is best described as a Bayesian learning model.

It is known that the weight K_t converges to its steady state value over time.

$$K = \frac{-[c + (1-\gamma^2)] + \sqrt{[c + (1-\gamma^2)]^2 + 4c\gamma^2}}{2\gamma^2}, \text{ where } c = q/h. \tag{8}$$

Once the weight converges to its steady state value, the estimate $\hat{\alpha}_t$ will change as an average of the prior estimate and new information. More specifically, when K converges to 1, $\hat{\alpha}_t$ will be equal to the current information (y_t); when K converges to

0, $\hat{\alpha}_t = \gamma\hat{\alpha}_{t-1}$. That is, when $K = 0$, $\hat{\alpha}_t$ becomes either constant (if $\gamma = 1$), or will approach 0 over time (if $\gamma < 1$). When $0 < K < 1$, $\hat{\alpha}_t$ will change as a weighted average of the prior estimate and new information.

We applied this model to candidate evaluation dynamics with the following assumptions: (1) voters evaluate candidates by estimating their unobservable qualities (α_t); (2) they see campaign information (y_t) about candidates as indicators of candidate qualities and update their evaluations $(\hat{\alpha}_t)$ upon receiving each new piece of campaign information; (3) they believe the true qualities of candidates change over time according to the process in Eq. 5; and (4) they follow the "optimal" updating rule given by Eq. 7.

This Bayesian model has five parameters: the initial estimate and the uncertainty associated with it ($\hat{\alpha}_0$ and P_0), a parameter that governs changes in $\hat{\alpha}_t$ over time (i.e., γ), and the variances of the normally distributed error terms, η_t and ε_t (i.e., q and h). Paralleling our JQP simulation, we determined the values of the five parameters in the following way. First, initial estimates of candidate evaluations $(\hat{\alpha}_0)$ were stochastically generated from the NAES 2000 pre-campaign survey just as in the JQP simulation. Second, we estimated the remaining four free parameters, choosing values that produced the best model fit (the highest correlation between actual and simulated evaluations over time), which turned out to be $P_0 = 4.3$, $\gamma = 1$, $h = 0.01$, and $q = 0$.[9] That is, the best model fit was obtained when Bayesian agents perceived that the true qualities of candidates do not change over time during the election ($\gamma = 1$) with no error ($q = 0$), when they believe that the campaign information reflects the true qualities of candidates with very small error ($h = 0.01$), and when they are relatively uncertain about their initial evaluations ($P_0 = 4.3$).

Since Bush and Gore were quite well-known to the public before the election campaign began, it may be the case that the survey respondents' evaluations had already reached stable state prior to our first observation in the NAES 2000. To incorporate this possibility, we ran a separate simulation assuming that the model had already converged to a steady state before the period of observation. In this second Bayesian simulation, the initial estimates of candidate evaluations $(\hat{\alpha}_0)$ were generated as before. The steady state value of the weight (K) that produced the best model fit was then chosen, which turned out to be $K = 0.56$.[10]

Bayesian Simulation Results

Table 4 shows the across-time correlations between simulated and actual evaluations when the Bayesian model was employed with and without assuming a steady state. Section A reports the distribution (mean and standard deviation) of across-time correlations, averaged across the two candidates, for the 100 simulations,

[9] The parameter search space was extensive; the values within the interval [0, 10] with stride 0.1 for P_0, h, and q, and those within [0, 1] with stride 0.01 for γ were first examined. Then, the values within [0, 1] with stride 0.01 for γ, h, and q were examined in a refined search. For very small values of h and q (0, 0.01, and 0.02), simulation results and model fits were almost identical. With the optimized parameter values, the weight, K_t, did not converge during the simulation.

[10] The values within [0, 1] with stride 0.01 were examined.

Table 4 Bayesian model fit—across-time correlations between survey and simulated evaluations

Ideological group	Not assuming steady state		Assuming steady state	
	Mean	Std. Dev.	Mean	Std. Dev.
A. Newsday/New York Times				
Strong liberal	0.406	0.28	0.119	0.2
Liberal	0.546	0.23	0.13	0.18
Moderate	0.554	0.1	0.328	0.11
Conservative	0.270	0.1	0.105	0.16
Strong conservative	0.142	0.1	0.05	0.21
All voters	0.622	0.09	0.427	0.09
B. Wall Street Journal				
Strong liberal	0.262	0.29	−0.008	0.19
Liberal	0.362	0.26	0.134	0.12
Moderate	0.443	0.1	0.179	0.11
Conservative	0.202	0.08	0.338	0.08
Strong conservative	0.323	0.15	0.557	0.1
All voters	0.525	0.08	0.361	0.09

broken down by ideological group, when the Bayesian agents read the liberal newspapers. Section B reports the same information when the stream of campaign information came from the *Wall Street Journal*. The first two columns report the distribution when the model was not assumed to have reached a steady state and the last two columns report the distribution when it was. Clearly the model provided a better fit to the NAES data when we did not assume it had reached steady state before the campaign began.

Although the Bayesian model did a reasonable job in tracking the observed changes in candidate evaluations over the course of campaign events, the Bayesian model fit was significantly lower than that obtained in the JQP simulations that we reported earlier. For instance, the mean correlations between the evaluations of all simulated agents and all survey respondents were 0.61 for the liberal media and 0.53 for the conservative media, as compared with 0.90 and 0.77, respectively, for the JQP simulations. Although not reported here, the across-group correlations in the Bayesian simulations were also substantially lower than those obtained for JQP for both media types. Though both models did reasonably well, JQP clearly outperformed the Bayesian model in tracking the observed changes in NAES candidate evaluations.

Most importantly, it turns out that the Bayesian learning model generates much more volatile changes in candidate evaluations over time than were observed for the survey respondents, while JQP produces changes of comparable magnitude to the observed. Table 5 compares the magnitude of change in candidate evaluations over time for each ideological group (measured by the standard deviations of the evaluations over time) in the JQP simulations, the Bayesian simulations, and for actual NAES respondents. The Bayesian agents' candidate evaluations were much

Table 5 Magnitude of changes in evaluations over time: JQP versus Bayesian learning model

	NAES data	JQP	Bayesian	Bayesian (steady state)
A. Liberal media (Newsday/New York Times)				
Strong liberal	5.27	2.71	13.83	16.34
Liberal	4.0	2.63	12.65	14.25
Moderate	2.39	2.06	12.08	12.62
Conservative	2.51	2.0	12.09	12.2
Strong conservative	2.39	2.1	13.0	14.8
All voters	2.24	1.95	11.82	11.91
B. Conservative media (Wall Street Journal)				
Strong liberal	5.27	2.28	13.11	18.47
Liberal	4.0	2.27	12.26	17.35
Moderate	2.39	1.99	11.34	16.52
Conservative	2.51	2.01	11.86	16.42
Strong conservative	2.39	2.18	13.3	18.04
All voters	2.24	1.95	11.21	16.02

more volatile in all ideological groups than either the survey respondents or JQP agents. Moreover, when a steady state is assumed, the Bayesian agents' evaluations are even more volatile.

The Bayesian agents' candidate evaluations moderated over the election campaign, while those of both real respondents and JQP agents polarized. Section A in Table 6 compares candidate evaluation differentials (Bush minus Gore) before

Table 6 Attitude polarization for liberal media: JQP versus Bayesian learning model

	Data		JQP		Bayesian	
	Before	After	Before	After	Before	After
A. Candidate differential before GOP convention versus after the debates						
Strong Liberal	−21.5	−37.0	−13.7	−20.1	−20.3	−10.6
Liberal	−14.8	−30.2	−10.4	−16.0	−15.2	−9.0
Moderate	−0.4	−4.9	−2.3	−4.4	0.1	−4.8
Conservative	25.4	29.3	12.6	14.8	25.4	1.9
Strong conservative	39.5	46.5	20.1	24.0	40.9	5.1
Average (absolute)	20.3	29.6	11.8	15.9	20.4	6.3
Initial evaluation	NAES data		JQP		Bayesian	
	Change	%	Change	%	Change	%
B. Average changes across candidates						
Preferred	4.5	6.55	2.13	3.22	−14.04	−20.38
Not preferred	−4.74	−9.48	−1.92	−3.52	0.03	0.06
Positive	1.74	2.7	0.85	1.33	−12.40	−18.98
Negative	−4.45	−10.3	−1.62	−3.12	5.57	12.86

the GOP convention and after the three debates. Where these differentials increase over the campaign, we see polarization; where they decrease, we see moderation. Across all ideological groups, the survey respondents' and JQP agents' evaluations polarized over the campaign, while the Bayesian model's evaluations significantly moderated. Section B breaks these patterns down for the initially preferred, not preferred, positively evaluated, and negatively evaluated candidates (note that even if both candidates are negatively or positively evaluated, one will be preferred). The first panel shows that both the survey respondents' and JQP agents' evaluations of initially favored candidates increased over the campaign (on average by 4.5 and 2.13 points on the 0–100 thermometer scale, respectively) and those of initially disfavored candidates decreased (on average by -4.74 and -1.92, respectively). By contrast, the Bayesian agents' evaluations of initially preferred candidates significantly decreased (on average by -14.04 points) and those of initially disfavored candidates stayed about the same. As the second panel shows, similar patterns were observed for positively and negatively evaluated candidates. We obtain qualitatively identical results for simulations using the conservative media source and when we assume a steady state for the Bayesian model.

Discussion: Responsiveness, Persistence, and Polarization

The NAES 2000 data show that evaluations of both Bush and Gore responded to the flow of campaign information, but within a relatively narrow range constrained by priors, and eventually preferences became more extreme. The empirical data show responsiveness, persistence, and polarization in candidate evaluations. JQP closely reproduces these observed dynamics, while the Bayesian learning model we tested accounts for neither persistence nor polarization for any ideological group, whether responding to liberal or conservative media. This difference in performance obtains despite the fact that both models start with the same initial knowledge structures and receive the same set of campaign information.

We do not argue that it is impossible to account for the persistence and responsiveness of political attitudes using a Bayesian learning model. In fact, in a very loose sense, JQP itself may be viewed as a Bayesian learning model in that prior beliefs play a critical role in learning in the model. On the other hand, like Bartels (2002) we do insist that any model that does not contain processing mechanisms capable of the differential subjective interpretation of information based on priors cannot account for the empirical dynamics of persistence, responsiveness, and polarization.

JQP captures all three qualities of the dynamics of candidate evaluation *because it models motivated reasoning*. It depreciates attitudinally contrary information but accepts congruent information more or less at face value. Unlike the Bayesian model, JQP is not a passive responder but a motivated skeptic in how it weighs new information. Specifically, in JQP the prior belief structure for an attitude object determines how incoming information will be perceived (through the patterns of activation in memory), and when attitudes are updated information that favors priors is weighted more heavily. JQP biases the processing of incoming information in the direction of prior attitudes—the farther away from the prior, the more likely this information will be challenged and discounted. Note that motivated reasoning is the direct consequence of the processing mechanisms that define JQP (and we believe *homo sapiens* as well).[11] By contrast, the naïve Bayesian learning model does not respond to incoming information on the basis of consistency with priors. It updates its prior evaluations to make them more accurate estimates of candidate qualities; it does not denigrate new information that challenges priors. For any set of campaign messages that contains both consistent and inconsistent information, the Bayesian model's candidate evaluations will tend to fluctuate with each new piece of information it processes, and it will likely moderate over time. Given the same set of mixed campaign messages, JQP's evaluations will be more persistent and preferences will tend to polarize over time.

As noted by Bartels (2002), real campaigns often provide precisely this mix of pro and con information that makes the difference between a passive learning model and an active motivated reasoner important. Just as in the media accounts we use to capture the flow of information in the 2000 presidential election, there is generally reason to both like and dislike all candidates. Motivated citizens impose order on an environment of contradictions, while passive responders are pushed one way and then the other. In this sense, motivated skepticism can provide healthy inertia for a democratic system. If skepticism becomes dogmatism, however, so that citizens are incapable of changing their minds even with compelling and consistent information against priors, motivated reasoning will be dysfunctional for democracy. In general, JQP is responsive to new information, though it takes a great deal of counterevidence to overturn very strong priors. Weaker priors, as are observed for many moderate agents in the JQP simulation, are more responsive to new information. This finding confirms earlier empirical evidence (Bartels 2002; Taber and Lodge 2006): partisans with strong priors are most biased in their processing of contemporary information, while nonpartisan reasoners respond to new information with less bias.

The discrepancy between a motivated reasoning model like JQP and a naïve learning model like Gerber and Green will depend on the nature of the information environment. A purely one-sided information environment (e.g., all conservative news), for example, will eventually lead both motivated reasoners and more passive learners to shift their attitudes in the direction of the information (though some motivated skeptics will initially resist). A segregated information environment that provides different but consistent information streams to liberals and conservatives will polarize skeptics and passive learners alike, though a motivated reasoning model could more easily explain how selective exposure could lead to such a segregation of citizens into separate "information publics." Finally, ideologically consistent campaigns would diminish the difference between motivated reasoning and passive learning, because they present information that is consistent with the priors of many citizens. For example, a campaign between a "pure" conservative and a "pure" liberal candidate may not challenge the priors of ideologically

[11] Formally, JQP will be a motivated reasoner as long as $\gamma > 0$, $0 < \delta < 1$, $0 < \rho < 1$ and its belief structure is reasonably consistent.

sophisticated citizens enough to motivate skepticism. By contrast, a campaign between less consistent candidates, who take stands that are sometimes liked and sometimes disliked by supporters and opponents, will maximize the differences between motivated reasoning models and passive learning models. This is important because there is little doubt that real campaigns vary considerably in the degree to which they challenge citizens' priors.

Several important implications come to the fore: First, these results suggest the centrality of motivated reasoning to the processes of candidate evaluation. A model that incorporates motivated mechanisms was able to reproduce the behavior of NAES survey respondents over time, while a passive learning model could not. Without motivated reasoning, it would be difficult if not impossible to provide a psychologically realistic account for why and how voters with opposing ideological structures could simultaneously maintain their candidate evaluations while responding to a common information stream. Second, JQP offers a solution to the controversy over the persistence and responsiveness of political attitudes. Motivated reasoning implies that our political attitudes are both persistent and responsive. That is, while we do respond to campaign information, our prior beliefs influence how we interpret incoming information and how we incorporate it into our attitudes, which are thereby resistant to fundamental change. Third, any learning model that does not incorporate motivated reasoning will have difficulty accounting for the polarization of candidate evaluation commonly observed over the course of political campaigns. Not only will citizens be skeptical about challenging information, but their attitudes will likely become more extreme as a result of this skepticism.

When priors are not challenged, citizens simply respond to campaign information, taking it at face value. Both JQP and the naïve learning model account for this behavior. When strong priors are challenged, however, citizens persist in their prior attitudes, and eventually may polarize. JQP accounts for this behavior while the naïve model does not. Moreover, JQP explains the psychological mechanisms that drive this behavior while also explaining a wide range of empirical regularities from social and cognitive psychology. This, we believe, is a significant step forward in our understanding of how citizens develop and change their political beliefs and attitudes.

Acknowledgments This research was supported by an NSF dissertation grant to Sung-youn Kim, research grants for Teragrid supercomputing resources (TG-SES050006 N, TG-SES050007 N), and an NSF research grant to Milton Lodge and Charles Taber (SES0241282).

References

Abelson, R. (1963). *Computer simulation of personality*. New York: Wiley.

Achen, C. (1992). Social psychology, demographic variables, and linear regression: Breaking the iron triangle in voting research. *Political Behavior, 14*, 195–211.

Anderson, N. (1968). Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology, 9*, 272–279.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of mind. *Psychological Review, 111*, 1036–1060.

Bartels, L. M. (2002). Beyond the running tally: Partisan bias in political perceptions. *Political Behavior, 24*, 117–150.

Bradley, M., & Lang, P. (1999). *Affective norms for English words (ANEW)*. Gainesville, FL: The NIMH Center for the Study of Emotion and Attention, University of Florida.

Bullock, J. G. (2009). Partisan bias and the Bayesian ideal in the study of public opinion. *Journal of Politics, 71*, 1109–1124.

Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. (1960). *The American voter*. New York: Wiley.

Downs, A. (1957). *An economic theory of democracy*. New York: Harper and Row.

Fazio, R. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion, 15*, 115–141.

Gerber, A., & Green, D. P. (1998). Rational learning and partisan attitudes. *American Journal of Political Science, 42*, 794–818.

Hamill, R., & Lodge, M. (1985). Cognitive consequences of political sophistication. In R. R. Lau & D. O. Sears (Eds.), *Political cognition* (pp. 69–93). Hillsdale, NJ: Lawrence Erlbaum.

Hastie, R., & Pennington, N. (1989). Notes on the distinction between memory-based and on-line judgment. In J. Bassili (Ed.), *On-line cognition in person perception* (pp. 1–19). Hillsdale, NJ: Lawrence Erlbaum.

Kim, S. (2005). A model of political information processing: The dynamics of candidate evaluation and strategic behaviors in social dilemma games. Unpublished Dissertation, Stony Brook University.

Kinder, D. R. (1998). Opinion and action in the realm of politics. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 778–867). Boston: McGraw-Hill.

LeDoux, J. (1996). *The emotional brain: The mysterious underpinnings of emotional life*. New York: Simon and Schuster.

LeDoux, J. (2003). *Synaptic self: How our brains become who we are*. London: Penguin Books.

Lodge, M., Steenbergen, M., & Brau, S. (1995). The responsive voter: Campaign information and the dynamics of candidate evaluation. *American Political Science Review, 89*, 309–326.

Lodge, M., & Taber, C. S. (2000). Three steps toward a theory of motivated political reasoning. In A. Lupia, M. McCubbins, & S. Popkin (Eds.), *Elements of reason* (pp. 183–213). New York: Cambridge University Press.

Lodge, M., & Taber, C. S. (2005). The primacy of affect for political candidates, groups, and issues: An experimental test of the hot cognition hypothesis. *Political Psychology, 26*, 333–487.

Miller, G. A. (1957). The magic number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–93.

Niemi, R. G., & Jennings, M. K. (1991). Issues and inheritance in the formation of party identification. *American Journal of Political Science, 35*, 970–988.

Page, B. I., & Shapiro, R. Y. (1992). *The rational public*. Chicago: University of Chicago Press.

Romer, D., Kenski, K., Waldman, P., Adasiewicz, C., & Jamieson, K. H. (2003). *Capturing campaign dynamics: The National Annenberg Election Survey*. Oxford: Oxford University Press.

Taber, C. S. (2003). Information processing and public opinion. In D. O. Sears, L. Huddy, & R. Jervis (Eds.), *Oxford handbook of political psychology* (pp. 433–476). Oxford: Oxford University Press.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50*, 755–769.

Taber, C. S., & Steenbergen, M. R. (1995). Computational experiments in electoral behavior. In M. Lodge & K. McGraw (Eds.), *Political judgment: Structure and process* (pp. 141–206). Ann Arbor, MI: University of Michigan Press.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of the survey response*. Cambridge: Cambridge University Press.

Zajonc, R. B. (1984). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35*, 151–175.

Zajonc, R. B. (2000). Feeling and thinking: Closing the debate over the independence of affect. In J. Forgas (Ed.), *Feeling and thinking: The role of affect in social cognition* (pp. 31–58). Cambridge: Cambridge University Press.

Zaller, J. R. (1992). *The nature and origins of mass opinion*. Cambridge: University of Cambridge Press.

Zaller, J., & Feldman, S. (1992). A simple theory of the survey response. *American Journal of Political Science, 36*, 579–616.