

Unconscious Thinking on Political Judgment, Reasoning, and Behavior

We are told by the astrophysicist Michio Kaku that 6.4 percent of the universe is visible, with another 23 percent unseen but measurable, leaving much of the universe in the dark. It is much the same in our inner world, where most thinking occurs outside of awareness, available to neither introspection nor direct observation. Humans are designed to process rapidly and *implicitly* enormous quantities of environmental and internal data. But our ability to focus *explicit* thought is severely limited. By and large, the social sciences are not well prepared to understand this duality of cognition, and political science is no exception. Grounded in an Enlightenment view of Rational Man, political science has been dominated by models of conscious control and deliberative democracy. Rational and intentional reasoning, in this conventional view, *causes* political behavior.

This is a book about unconscious thinking and its influence on political attitudes and behavior. It is a book about powerful affective and cognitive forces that motivate and direct deliberation and political action outside of conscious awareness and control. It is a book about rationalizing, rather than rational, citizens.

What people think, feel, say, and do is a direct function of the information that is momentarily accessible from memory – be it the recall of facts and feelings, the recollection of experiences, or the turning of goals into action. Political behavior and attitudes are very much a function of the unconscious mechanisms that govern memory accessibility. But we political scientists know very little about the processes that underwrite individual variation in beliefs and behavior. We know about variation in public opinion as indicated by verbal self reports. We routinely ask respondents for their party and candidate preferences, their approval of policy proposals, and how warmly they feel toward one or another group, and we are often able to relate these explicit measures through sophisticated multivariate analyses that we interpret as revealing

underlying causal processes. There has also been considerable growth in the use of controlled experiments to determine causality, but most of these also rely on overt verbal responses that may not reveal an underlying implicit process. This reliance on direct, explicit measures of political beliefs and attitudes is intensely problematic, assuming as it does that people have accessible beliefs and attitudes, that they are willing and able to voice them, and that these self-reports are causally related to their political behaviors.

Though it has gone largely unnoticed in political science, we are witnessing a revolution in thinking about thinking. Three decades of research in the cognitive sciences, backed by hundreds of well-crafted behavioral studies in social psychology and now evidence from the neurosciences, posit *affect-driven, dual-process modes* of thinking and reasoning that directly challenge the way we political scientists think about, measure, and interpret political beliefs and attitudes. Central to such dual-process models is the distinction between the unconscious (“System 1,” “implicit”) and conscious (“System 2,” “explicit”) processing of judgments, preferences, and decisions. System 1 processes are spontaneous, fast, effortless, and operate below conscious awareness, whereas System 2 processes are slow, deliberative, effortful, and self-aware.

Given the serious real-time limitations of conscious processing, we humans have evolved compensatory heuristics, including a System 1 *likeability heuristic* that automatically links positive and/or negative affect to familiar social objects in long-term memory. Once associated, this felt positivity or negativity strongly influences downstream thinking and reasoning. What especially attracts our interest as political scientists to such dual-process models is the finding that unconscious processes are continually at work, with effects that appear to be most influential when the most knowledgeable among us think hard about an issue and carefully weigh the pros and cons when forming opinions and making choices.

The Ubiquity of Unconscious Thinking

Cognitive scientists estimate that the human capacity for processing sensory experience is about 11 million bits per second (Norretranders, 1998). The visual system takes up about 90 percent of this total capacity, processing roughly 10 million bits of visual information per second. No more than 40 bits per second of this visual information enters conscious working memory, so we become aware of only 1/250,000 of what we see! Similarly, a healthy human brain processes 1 million bits of tactile information and 100,000 bits of auditory information, while we at best become aware of just 5 bits of tactile and 30 bits of auditory information per second. When we read (with or without moving our lips) we process a maximum of 45 bits per second. More limited still is our capacity to consciously think and reason, where we are able to keep in the focus of attention only about 7 ± 2 chunks of information (Miller, 1956).

About 98 percent of what we experience, our very connection to the outside world, are whispers that come and go unnoticed.

What are the consequences of this colossal difference between conscious and unconscious experiences for thought and action? What types of information activate unconsciously when citizens watch a candidate debate, see a campaign ad, argue politics with friends, ruminate about a political issue, answer a pollster's question, or enter the voting booth? Where, when, and why will conscious and unconscious processes reinforce one another? What happens when unconscious influences are at odds with conscious control? When and how can unconscious influences be overridden (Bodenhausen and Todd, 2010)?

Research across the cognitive and neurosciences demonstrates the profound impact of unconscious processing on the content of our thoughts, how we reason, and consequently the choices we make (Ferguson and Porter, 2010; Hassin, Uleman, and Bargh, 2005; Perugini, Richetin, and Zogmaister, 2010). To place this empirical literature in perspective, and reassure readers that the “unconscious” explored here and in the contemporary psychological literature is not the subterranean id, ego, or superego of Freud, or the psychoanalytic analyses popular in the mid-twentieth century (Erikson, 1950; George and George, 1956; Lasswell, 1930), let us operationalize the unconscious in terms of objective and subjective thresholds of perception.

An objective threshold, as can be measured by brain-wave patterns, must be passed for an external stimulus event to enter one of the sensory systems. A subjective threshold is passed if the stimulus event enters conscious awareness. There are three possibilities:

- If the objective threshold is not passed, perception does not occur and there is no registration of the event on the senses. Essentially, a nonevent with no impact on information processing.
- If the objective threshold is passed but the subjective is not, we have unconscious perception — a sensory experience passes objective thresholds without ever entering conscious awareness. Such *Consciously Unnoticed Events* (Type 1 CUEs or interchangeably called Type 1 primes) escape notice; seen, registered, but consciously unnoticed. An objectively perceived stimulus may not reach conscious awareness for many reasons: because it occurred too rapidly or too peripherally to be noticed, or one is momentarily distracted.
- If the subjective threshold is passed, we have explicit conscious perception, the stuff of everyday experience. But — this very common — we may “see” the stimulus without realizing its influence on our thoughts, feelings, preferences, and choices. For such *Consciously Unappreciated Events* (Type 2 CUEs or interchangeably Type 2 Primes), the individual is consciously aware of the stimulus, say the American flag in the background of a candidate's speech, but its impact on thought, reasoning, and choice is not seen as being influential.

Unconscious primes are ubiquitous in the real world (Bargh, 1997), the playthings of advertisers selling detergents and presidential candidates, where the men and women in beer and car commercials are unusually attractive and fun loving; the smokers in cigarette ads look preternaturally healthy; the men tout-ing erectile dysfunction medications appear uncommonly virile. Laugh tracks in situational TV comedies, although widely bemoaned, nonetheless enhance audience enjoyment. Worse yet, all types of humor, whether real or feigned, are commonly used to mask deceptive advertising (Shabbir and Thwaites, 2007). And as we will show in multiple experimental demonstrations, such “incidental,” more-often-than-not diagnostically irrelevant Type 1 and Type 2 primes prove to be powerful influences on how people think about and evaluate political leaders, groups, and issues.

Unconscious events and processes can drive political behavior in two ways: they may directly trigger a snap judgment or response entirely out of awareness, or they may indirectly drive behavior through their influence on conscious thought processes. A great deal of psychological research has demonstrated the direct causal process, but there has been comparatively little research on the mediated impact of implicit processes.

Implicit Cues in the Real World and in the Laboratory

Because citizens are confronted with more information than they can consciously handle, it should come as no surprise that they take mental shortcuts to arrive at their vote decisions, including endorsements, opinion polls, physical attractiveness, elite opinion, and feelings toward social groups (Mondak, 1994) – and of course party identification (Bartels, 2000; Goren et al. 2009; Jackman and Sniderman, 2002; Lau and Redlawsk, 2006; Riggle et al., 1992; Sniderman, 2000; Sniderman, Brody, and Tetlock, 1991). Reliance on one or another heuristic seems a reasonable strategy to the extent that it helps align a candidate’s issue positions and attributes with the voter’s interests and values (Lau and Redlawsk, 2006) or more generally improves the quality of decisions (Kahneman, Slovic, and Tversky, 1982).

But we believe and hope to demonstrate another, even faster, more readily available and general heuristic exists that may provide quicker and “better” candidate evaluations: a System 1 *likeability heuristic* stored as an implicit attitude unconsciously guides preferences in accord with the citizen’s history of information processing. Implicit attitudes or feelings about individuals, social groups, and ideas can exist outside of subjective awareness, affective tallies capture the evaluative implications of prior conscious and unconscious thinking about these objects, and these feelings come spontaneously to mind when their associated objects become targets of thought.

A great deal of psychological research shows the impact of implicit attitudes on a variety of social behaviors (Gawronski and Payne, 2010; Petty, Fazio, and Briñol, 2009), though the relationships among implicit and explicit attitudes

remain controversial (De Houwer, 2009). For example, implicit racial attitudes have been repeatedly shown to influence social behaviors, though they often diverge from explicit self-report measures of racial attitudes (Dovidio et al., 2009; Greenwald and Nosek, 2009; Nosek and Smyth, 2007). We believe that it would be a serious error to make a too-sharp distinction between implicit and explicit attitudes and we resist doing so (Sherman, 2009). Our view is that implicit and explicit attitudes are different responses from a single underlying memory system. Explicit attitudes are consciously considered responses for which one has the time and motivation to form a response. They will be influenced by myriad unnoticed factors, but somewhere in the decision stream will be an opportunity for control and consciously reasoned thought. Implicit attitudes are affective responses to stimuli that one cannot control or consciously reason about. It is more likely that an implicit response reflects affect stored directly with a memory object (what has been called an online tag in the research literature), but these too will be influenced by extraneous factors. It is a mistake to think of one as more “true” than another, and both are subject to bias, though of a different kind.

Is it possible to like someone or something without any conscious awareness of how or why this preference came to be? In his presidential address to the American Psychological Association, Robert Zajonc (1980) provides a simple experimental example for how “Preferences Need No Inferences.” A sample of non-Chinese Americans were briefly shown a number of Chinese ideographs and later asked to evaluate how aesthetically pleasing they were. The ideographs were shown zero, one, two, or three times, though participants were not aware of the multiple exposures and could not later identify which characters in a test set had been presented to them. Nevertheless, the more often they were shown a symbol the more they found it pleasing, a finding labeled the “mere exposure effect.” Preferences were altered without the objects even being recognized. In a final definitive demonstration that the mere exposure effect operates unconsciously, Murphy and Zajonc (1993) replicated the study using subliminal exposures to the ideographs (i.e., presentations too rapid for conscious perception).

Mere exposure can also influence other types of social judgments. Jacoby, Kelley, Brown, and Jasechko (1988) found that judgments of whether a name is that of a famous person (i.e., Is Sebastian Weisdorf famous?) are influenced by previous exposure to the name, even when it was presented on a list explicitly labeled Nonfamous People. Names were accurately judged to be nonfamous immediately after exposure to the list, but twenty-four hours later as recall of the source of information faded from memory, the residue memory trace was sufficient for many of those on the list to become famous overnight. Mere exposure, bolstered by this sleeper effect, changed the accessibility of names, making them appear more familiar and hence mistakenly identified as famous. This effect mimics what is routinely found in studies of persuasion where familiar arguments are judged more believable (Eagly and Chaiken, 1993),

where in advertising repetition builds brand name identification (Warshaw and Davis, 1985), and where candidate name recognition is, after money, the most critical step in winning an election (Kleinnijenhuis, van Hoof, and Oegema, 2006). Here again, conscious and unconscious processing may go their separate ways.

Unconsciously processed cues operating in the political realm can impact the evaluations of known candidates and their electoral success. The 1960 Nixon-Kennedy preelection debate is a well-known political example of noticed-but-unappreciated effects: seventy million people watched the first televised presidential debates in American history between Richard Nixon and John Kennedy. Nixon, recently out of the hospital, refused make-up; Kennedy had been campaigning in California and had the tan to show for it. Television viewers, apparently distracted by Nixon's pallid look and five-o'clock shadow, thought Nixon shifty and untrustworthy, while radio listeners, who had little to go on but the substance of the debates, thought Nixon the clear winner. The familiar version of this story is used to illustrate how image can dominate substance in politics; in our terms, how System 1 implicit processing can lead voters astray from the solid moorings of conscious deliberation. But as Malcolm Gladwell (2005) points out, the familiar version of the story has it backwards: Nixon did indeed turn out to be shifty and untrustworthy. Viewers' implicit, affective responses to the candidates' appearances proved to be more accurate than judgments based presumably on a less-biased, more careful consideration of issue positions and policies.

Similarly, facial expressions of news broadcasters influence the political judgments of viewers. In coverage of the 1976 presidential election campaign, Friedman, DiMatteo, and Mertz (1980) found discernable differences in the perceived positivity of broadcasters' facial expressions when they uttered different candidates' names. Mullen and colleagues (1986) replicated this result with the 1984 presidential election and demonstrated further that a broadcaster's facial expressions influenced voters' political preferences. Specifically, voters came to favor the candidate for whom the broadcaster exhibited more positive facial expressions. The same effect in a different modality: Gregory and Gallagher (2002), analyzing the voice frequencies of candidates in nineteen nationally televised American presidential debates, found that this auditory cue signaled a candidate's relative social dominance within a debate and predicted his vote share in the election. Media effects without message – more accurately, media effects through implicit rather than explicit channels of communication.

Babad (1999, 2005) obtained similar noticed-but-unappreciated results in the domain of political interviews. She found, not only that TV newscast interviewers exhibited differential levels of positive and negative nonverbal behaviors toward the politicians they were interviewing, but that an interviewer's nonverbal behavior impacted the viewers' perceptions of the politician. In particular, a politician's image suffered when the interviewer appeared hostile rather than friendly.

Here is an even more subtle effect of an unappreciated cue on choice: Berger, Meridith, and Wheeler (2008) showed that budgetary support for education varied as a function of where people voted – whether in schools, churches, or firehouses – with voters more likely to favor raising state taxes to support education when voting in schools, even controlling for their political views. Clearly, the voters knew what building they were in but were not consciously aware of its influence on their vote choice. Ballot order effects provide another political case in point, where being listed first increased the vote count for 80 percent of candidates (Schneider, Krosnick, Ofir, Milligan, and Tahk, 2008).

Some cues seem so obvious it is hard to imagine an implicit effect, but the inference is nevertheless made unconsciously. Race messages in campaign advertising, for example, are more effective when they remain covert. Tali Mendelberg demonstrates this effect in *The Race Card* (2001) via an experimental analysis of the infamous Willie Horton campaign ads, in which presidential candidate Michael Dukakis used pictures and sounds to implicitly associate African Americans with crime with. When the race cues are made fully explicit in Mendelberg's study (that is, when subjects are alerted to their presence) they lose their power to influence political judgments. Another case in point was a 2004 MoveOn.org TV ad that showed images of Hitler before a photo of Bush raising his hand to take the oath of office, accompanied by the voice over, "A nation warped by lies. Lies fuel fear. Fear fuels aggression. Invasion. Occupation. What were war crimes in 1945 is foreign policy in 2003." Republican groups and Jewish organizations expressed outrage over the ad, which was quickly removed from the MoveOn.org website. Research suggests, however, that subtle propaganda would be more effective; an implicit message more powerful still.

In the mid-1990s, Mayor Rudolph Giuliani of New York City adopted a "quality of life" campaign fashioned on James Q. Wilson and George Kelling's (1996) "broken windows theory." In this theory, signs of disorderly and petty criminal behavior signal neighborhood decay and deterioration, which trigger more disorderly and petty criminal behavior. Giuliani's change in policy had more cops walking beats, city work crews painting over graffiti, sweeping streets and cleaning subways, towing abandoned cars, ticketing jaywalkers, punishing vandals, and rousting the homeless from city streets and parks. After the introduction of the campaign, petty crime rates in New York City dropped dramatically and polls showed an uptick in perceived quality of city life (which became a major talking point for Giuliani's later political campaigns). A change in policy that was essentially cosmetic eventually had real effects on the compliance behavior of citizens, in our interpretation because of the replacement of implicit cues of neighborhood decay with cues of orderliness and civic control.

Political judgments can be directly affected by irrelevant, nonpolitical cues as well. While theories of retrospective voting suggest voters should reward or punish incumbents for the things they can control (in particular, wars and the

economy), it is hard to imagine why voters should hold politicians accountable for such “acts of God” as earthquakes or floods. And yet in their analysis of retrospective voting in Woodrow Wilson’s 1916 reelection, Achen and Bartels (2006) find that a string of shark attacks in the summer months before the 1916 election cost Wilson about ten percentage points in New Jersey beach communities, with no effect inland. Closer to home is the Healy, Malhotra, and Mo (2010) finding that local college basketball and football wins impacted the vote for Obama. Such findings are hard to square with conventional normative models of conscious deliberation, but are compatible with the implicit effects of affective cues on candidate preference.

A major area of research pointing to robust effects of unconscious influences on snap judgments is the effect of facial attractiveness on evaluations, attitudes, and behavior. Here, as in the stereotypic inferencing of traits from gender, age, and race, the face is rapidly registered and spontaneously triggers stereotypic assumptions about the individual’s character, attitudes, and behavior. Three large meta-analyses covering more than 1,000 peer-reviewed psychological studies of physical attractiveness confirm significant experimental and correlational effects on a broad range of social attitudes and behaviors (Eagly, Ashmore, Makhijini, and Longo, 1991; Feingold, 1992; Langlois, Kalakanis, Rubenstein, Larson, Hallam, and Smoot, 2000). Whether a person is seen as attractive or unattractive, assumptions are brought into play. Across cultures, what is beautiful is assumed to be good, and all manner of negative traits may be attributed to those less physically blessed. As Langlois and colleagues point out, this research shows that implicit responses debunk the descriptive if not the normative validity of three popular folk maxims:

Whereas it is said that *beauty is in the eye of the beholder*, the empirical evidence shows widespread consensus as to who is or is not attractive, with correlations suggesting near unanimity: within culture, $r = .90$; across ethnic groups, $r = .88$; and across cultures, $r = .94$. Such levels of agreement support the probability of rather uniform implicit responses to the appearances of political candidates or opinion leaders.

While we are admonished to *never judge a book by its cover*, hundreds of studies report stereotypical attributions advantaging attractive children in school and adults in their everyday lives and careers. It is routinely found that physical appearance exerts a strong influence on character perception, with scores of studies reporting a “beautiful-is-good” halo effect. The meta-analyses document that physically attractive people are perceived to be more sociable, dominant, extraverted, popular, and warm. Even among strangers a one second glance is enough to trigger an inference that an attractive man is more interesting, successful, intelligent, and virtuous. Strong correlations between attractiveness and particular attitudinal and behavioral characteristics have been found across cultures for both adults and young children, implying that a large part of this beauty-is-good projection effect is inborn and supplemented by nurture (Rhodes, 2006).

In general, a mere glance at an attractive face promotes a one-half standard deviation enhancement on positive personality traits, with about 64 percent of attractive people but only 36 percent of less attractive people perceived as having a better-than-average personality, the attractive seen as being more socially competent (70 percent vs. 30 percent), more worthy of attention (74 percent vs. 26 percent), more successful (68 percent vs. 32 percent), and if in need more likely to receive help (59 percent vs. 41 percent). Even in death the attractive are “advantaged,” their demise judged more tragic (Callan, Powell, and Ellard, 2007).

Finally, if it were true that *beauty is only skin deep*, there would not be a robust influence of self-rated attractiveness on measures of popularity, sociability, or objective measures of mental health. Physically attractive individuals have more sexual partners, find better-looking mates, become more professionally successful, make more than their fair share of decisions, and are happier than those of us below the median of physical good looks (Dion, Walster, and Berscheid, 1972). This “beauty premium” has been shown by Biddle and Hamermesh (1998) to positively impact attorneys’ wages, and – this unimaginable for elected office to political science associations – good-looking scholars are more likely to be voted into leadership positions of the *American Economics Association*.

The impact of physical appearance extends beyond attractiveness. A study by Mueller and Mazur (1996) found that ratings of facial dominance of West Point cadets (rectangular face, strong brow, square jaw) predicted later military rank. A follow up study (Little, Burriss, Jones, and Roberts, 2007) graphically manipulated facial dominance of alleged politicians and found that facial dominance affects voting decisions. Moreover, changing the context from peacetime to wartime promoted an even larger advantage for the dominant candidate.

What is important here is that physical appearance is registered but its inferential impact on character perceptions, evaluations, and behavior remains covert for those making the judgments. When this influence is pointed out, it is routinely denied. Given that facial appearance is one of the very first things we see in another person and that there are specific brain structures designed to detect and characterize faces, it is not surprising that attractive people prompt positive attributions which, entering the evaluation early, anchor and bias subsequent evaluations. Routinely, humans make positive attributions to attractive people without consciously realizing it, yet the magnitude of these effects is roughly the same as other variables in the social sciences (Eagly, 1996).

“Beautiful-is-good” stereotyping is alive in the political domain as well, where many of the same effects of attractiveness on snap judgments found in nonpolitical domains are matched in impressions of politicians, with attractive candidates seen as possessing more integrity, competence, likeability, and being better suited for public office (Rosenberg et al., 1986). For example, a

large-scale study of the 2003 parliamentary and 2004 municipal elections in Finland collected ratings by more than 10,000 web-survey respondents on a host of dispositional traits for a total of 1,900 facial photos of real political candidates. The finding: a one standard deviation increase in attractiveness was associated with a 20 percent increase in the number of votes over the average nonincumbent (Berggren, Jordahl, and Poutvaara, 2010). Similarly, in a study of the 2004 Australian election, where voting is compulsory and voters are handed a “How to Vote” card with pictures of the candidates, the more attractive of the two was associated with a 1.5 percent to 2 percent change in vote share, with this effect even larger in electorates with a higher share of apathetic voters (King and Leigh, 2010). Rosar, Klein, and Beckers (2008) found the same result for the state-wide elections in the largest German Bundesland, North Rhine-Westphalia, where campaign posters feature pictures of the candidates: attractive candidates – especially when their opponents are unattractive – garnered not only a larger vote share but also an increase in turnout.

While most of these studies have experimental participants view photos at their leisure in a contextually relevant frame, a great deal of information in addition to facial attractiveness can be gleaned in the blink of an eye (Gladwell, 2005). Here’s an “experiment” to try. On the next page are side-by-side photos of a pair of adult males, both candidates for the U.S. Senate (Figure 1.1). Turn the page, take no more than one second to scan the photos and return here.

Now which of the two candidates would you say is more competent?

In an important series of experiments reported in *Science*, Alex Todorov and his colleagues (2005; see also Olivola and Todorov, 2010) demonstrated that competence ratings based on a one-second exposure to paired photos of competing candidates predicted the 2004 House and Senate election outcomes at significantly better than chance levels (67.7 percent and 68.8 percent, respectively). Competence in the Todorov studies is modeled as a direct predictor of vote choice, and ratings were made of *unfamiliar* candidates by *naive* experimental participants *before* the 2004 congressional elections and the predictions are to the *actual* electoral outcomes, not vote intention. In other analyses, in addition to making competence judgments, participants evaluated the paired candidates on attractiveness, likeability, trustworthiness, and other dispositional judgments, all well-known to be important in the evaluation of political candidates (Kinder, Peters, Abelson, and Fiske, 1980; Funk, 1999). Now postdicting the 2000 and 2002 Senate races, Todorov and colleagues found what is also true in the National Election Studies: competence trumps the other trait assessments in accurately discriminating winners from losers. The inescapable implication of this research is that people can make substantively important attributions on a mere one second exposure to the facial photos of unfamiliar political candidates, and what is more, these snap judgments (typically taking little more than one second) discriminate winners from losers without any information or contextual cue other than being told the photos were of politicians. All this predictive power without party identification, ideological



FIGURE 1.1. A Pair of Senate Candidates from Todorov and Colleagues (2005)

proximity, or any of the traditional predictors of vote choice! Of course, it is possible that these more traditional levers of political judgment would be as or more influential on vote choice if they were available for respondents in these studies. But this fact does not overturn the importance of the finding that mere exposure to faces is sufficient to generate snap trait judgments and thereby alter vote choice.

A number of additional studies have replicated the general finding that appearance-based competence judgments predict election outcomes, while ruling out the alternative hypothesis that competence judgments simply reflect media-induced familiarity with the politicians. Lenz and Lawson (2007) asked American participants to make facial competence judgments of Mexican politicians. Their judgments predicted Mexican election outcomes and accounted for 18 percent of the variance in vote shares, though these participants were never exposed to the Mexican media. Experiments by Antonakis and Dalgas (2009) are especially revealing here because they address the possible confound between competence and incumbency and raise the question as to how facial appearance predicts vote choice. Judgments collected from a sample of 1,106 Swiss adults predicted the winner and runner-up from the run-off stages of the 2002 French parliamentary elections and their competence ratings predicted the margin of victory.

Antonakis and Dalgas pushed the research question deeper by asking 681 children aged 5 to 13 years to play a computer game simulating a voyage on a difficult seagoing mission in which they chose which person (from the paired

photos of French parliamentarians) they would want to captain the boat from Troy to Athens. The premise for this study dates back to Plato's *Republic* (2000: 153): "Imagine then a fleet or a ship in which there is a captain who is taller and stronger than any of the crew, but he is a little deaf and has a similar infirmity in sight, and his knowledge of navigation is not much better." Plato argues that the crew (voters) cannot select a competent captain (ruler) because the crew is beguiled by appearances. The children in Antonakis and Dalgas's experiment (mean age 10.3 years) predicted the French election outcome from their choice of ship captain with a correlation 0.71, which was indistinguishable from the adults' predictive success. These findings tell us that appearance-based trait inferences develop quite early and are surprisingly stable across age cohorts. Whatever the underlying process, both children and adults use facial cues rather than any in-depth processing.

Let's take the process one level deeper than cognitive deliberation can fathom. Social scientists may find it hard to believe but there are many experiments in developmental psychology that show the effects of attractiveness on infants and toddlers younger than the adolescents engaged in the sea-faring adventure of Antonakis and Dalgas (see Pascalis and Slater, 2003). Because infants cannot tell you what they find attractive or tell you much of anything, researchers use a "preferential-looking technique" in which two faces are shown side by side for ten-second exposures while a video camera records the time the infant spends gazing at each of the pictures. The consensual assumption is that the longer the fixation the more the infant is attracted to, or ostensibly "likes" the face. In one of many such experiments, Langlois and colleagues (1987) showed 6-month-olds images of female faces previously rated by college students as more to less attractive. For each pairing of faces (none were "drop-dead gorgeous" or "grotesque"), they found that the infants fixed their gaze longer on the more attractive face. Pushing the paradigm to its limits, the Langlois team (1991) next examined the preferences of 3-month-old infants to four types of faces – Black men and women, White men and women – all previously rated on attractiveness. Results confirm earlier, less-well controlled studies, in showing that preference for attractive faces holds across genders and races.

But what is it about the faces of politicians that causes people to perceive the winners as more competent than the losers? From our viewing of C-Span it is certainly not the case that the real-world competence or intelligence of politicians is reliably related to facial appearance. Perhaps there is a negative relationship. Todorov and colleagues (as well as other experiments from here and abroad) show that attractiveness and age, along with competence are proximate predictors of vote choice, but they do not rule out the possibility that competence simply mediates the causal effect of attractiveness and age on vote choice. Working from the "beautiful-is-good" literature, Verhulst, Lodge, and Lavine (2010) reconsidered the Todorov (Todorov et al., 2005; Olivola and Todorov, 2010) analyses to test the hypothesis that competence ratings are

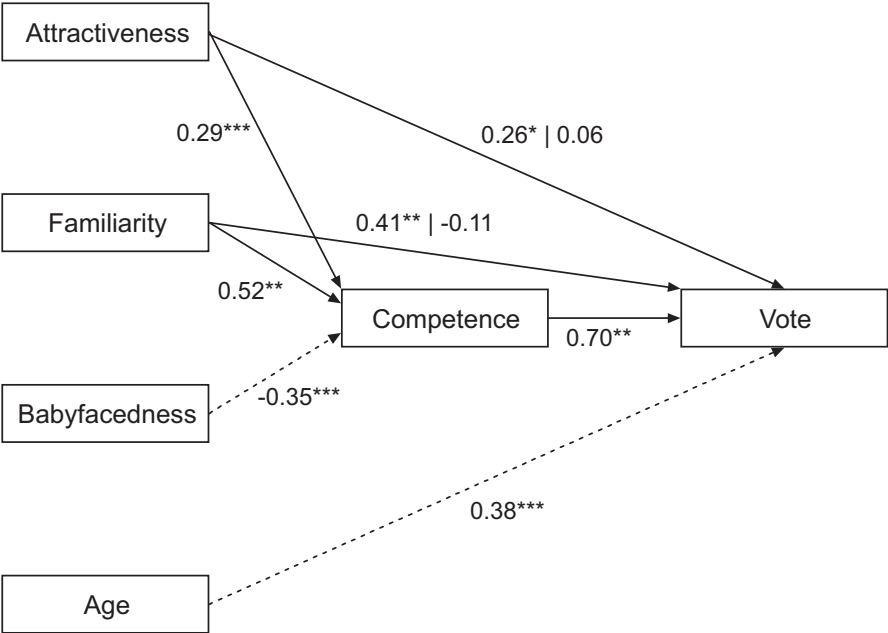


FIGURE 1.2. A Mediation Model of the Todorov and Colleagues (2005) Data

themselves derived from perceptions of facial attractiveness (as well as several other theoretically prior trait attributions).

Figure 1.2 reports the Todorov findings, rearranged into a mediational causal analysis to explain vote choice with four independent variables, three of which are mediated through competence attributions. Following the traditional mediational logic of Baron and Kenny (1986), Figure 1.2 shows three separate stages of regression analyses: first, we report the unmediated effect of each of the independent variables on vote choice, finding that attractiveness, familiarity, and perceived age all have a significant effect on vote choice, while babyfacedness does not have a direct effect; second, we report the effect of each independent variable on the mediator, finding that attractiveness, familiarity, and babyfacedness all significantly predict attributions of competence; finally, we report the effect of the mediator on the dependent variable while controlling for all four independent variables, finding that competence is the strongest predictor of vote choice, while the direct effects of attractiveness and familiarity drop out (the second coefficients reported for those paths in Figure 1.2). In short, Todorov’s data show that the causal pathways from attractiveness and familiarity to vote choice travel indirectly through the more proximate, causally later assessments of competence. In fact, 70 percent of the effect of attractiveness on vote choice, and 89 percent of the total effect of familiarity is mediated

through competence. Perceptions of candidate age exert a direct causal influence on vote choice without any indirect effect through competence, while babyfacedness has only an indirect influence.

Judgments of competence are clearly related to vote choice as Todorov and colleagues suggest (and as is shown repeatedly in the National Election Surveys), but the spontaneous process of making competence judgments appears to be preceded by an even earlier automatic assessment of attractiveness and familiarity. Given the emerging consensus that judgments of attractiveness have a biological basis, with specific brain structures engaged in the recognition of faces and facial expressions (Ekman, 2007), it is not surprising that these thin-sliced, one-second evaluations of political candidates are influenced by an even more primary evaluation of attractiveness. In addition to predicting higher levels of competence, physical attractiveness of politicians significantly predicts higher levels of likeability, integrity, and trust, all of which have also been repeatedly linked to the evaluation of political candidates and vote choice (Kinder, Peters, Abelson, and Fiske, 1980).

A cautionary note: neither we nor Todorov claim that the momentary effects of attractiveness on vote choice trump incumbency, party identification, issue proximity, or the many other factors known to predict congressional elections. Nor is anyone arguing that this bias cannot be corrected (Hart, Ottati, and Krumdick, 2011), although not easily, requiring as it does the conjunction of cognitive capacity to recognize the influence of physical attractiveness on one's judgment, the belief that the bias is inappropriate, and the motivation to correct the evaluation downward for an attractive candidate and upward for an unattractive contender. Rather, the point is that a simple glance generates inferences that have political import. Not surprisingly, a Todorov-like study by Atkinson, Enos, and Hill (2009) shows that political parties running candidates in competitive congressional elections selectively choose challengers with "higher quality faces." Across the ninety-nine Senate elections the authors found a significant "face quality" effect for both independent and partisan voters, but no instance where face effects in competitive elections changed the electoral outcome.

The question asked for millennia but still a puzzle today is why we are predisposed to find attractive faces so interesting (it cannot be a familiarity or socialization effect) and why preschoolers, youngsters, teenagers, and adults go beyond attractiveness to infer "beauty is good" given that these inferences appear not to facilitate accurate social judgments. One possibility consistent with the existing empirical evidence is that such inferences are based on cues that have adaptive significance (Todorov et al., 2008; Zebrowitz, 2004; Zebrowitz and Montepare, 2008). There is a dark side to the attractiveness-competence relationship, of course, in that the intelligence of adults cannot be predicted from facial appearance (Zebrowitz, Hall, Murphy, and Rhodes, 2002), and – this is admittedly a leap of faith – some politicians may actually be more competent than others.

Nonverbal cues have impact even in situations where decisions are made thoughtfully with due deliberation: Zebrowitz and McDonald (1991), for example, found judicial decisions to be influenced by the facial features of defendants and plaintiffs: mature-looking defendants were required to pay larger penalties in small claims courts when the plaintiffs were babyfaced. The robust effects of attractiveness on perception and behavior lend credence to Blaise Pascal's claim in his *Pensees* (1660; 2010): "Cleopatra's nose, had it been shorter, the whole face of the world would have been changed" (180).

In addition to unconscious trait attributions and the pronounced halo effects of attractiveness, there are countless examples of even more "incidental" influences on political information processing. Here is a perfect example of what we see as a not-so-subtle attempt to manipulate political inferences. In a televised, thirty-second, 2007 Christmas message by presidential candidate Michael Huckabee to Iowans a week before the caucuses. A single frame of this campaign ad is presented on the next page (Figure 1.3). Glance at it quickly, and then come back here.

Did you notice the bookcase over Hucklebee's right shoulder? Did the bright white separators of the bookcase form a cross? Note that the bookcase/cross may or may not be noticed. Perhaps the bookcase-as-cross would be more likely noticed by evangelicals and register as positive, while for others the implications might be negative, perhaps seen as a right-cross jab at Mitt Romney's square jaw or a poke at his Mormon religion. There is also the possibility that the symbol would escape conscious awareness, but be registered unconsciously, and thereby not be open to critical appraisal.

Such "incidental" priming is of course commonplace in the world of commercial and campaign advertising and given the research demonstrating that even brief exposures can impact preferences, it was to be expected that "thin-sliced" exposures much too fast to be reliably noticed would find their way into advertising as "hidden persuaders" and then into the selling of the president.

In his prophetic novel 1984, George Orwell (1949/2003) foretold of a future in which our thoughts, attitudes, and behaviors would be controlled by government-directed media. This prophecy gained plausibility in the late 1950s after the advertising executive James Vicary reported significant increases in Coke and popcorn sales after flashing three-hundredth-of-a-second directives to "Drink Coke" and "Eat Popcorn" during a movie. The results seemed staggering: movie sales of Coke and popcorn increased 18 percent and 58 percent, respectively. People were understandably appalled at this insidious mind-control technique. If it could be used to persuade people to buy snacks and soft drinks, what other behaviors might be subliminally manipulated? There is a problem with the results of the study, however: it never actually took place. Vicary made it up as a publicity stunt to generate interest in his struggling advertising agency. Hoax or not, most people are fearful of the possibility of being influenced by subliminal messages (Wilson and Brekke, 1994), and many countries prohibit it in advertising.



FIGURE 1.3. Huckabee Campaign Ad Image (2007)

Of course, the fact that Vicary's claim was a hoax did not establish that subliminal messages do not influence attitudes. Karremans, Stroebe, and Claus (2006) conducted two experiments to examine whether subliminal priming of a drink can affect people's choices for the brand, and, importantly, whether this effect is moderated by individuals' feelings of thirst. Both studies demonstrated that subliminal priming of a brand name (here, Lipton Iced Tea) positively affected participants' choice for, and their intention to drink the primed brand, but only for participants who were already thirsty. "You can lead a horse to water but. . . ."

As any self-respecting free marketer would predict, the priming of hidden persuaders would find its way into the selling of the president. In the 2000 presidential election campaign, the Republican National Committee aired a TV ad nationwide attacking Gore's prescription drug plan 4,400 times, costing the RNC \$2,576,000. When the final segment of the ad is run in slow motion, we can see the word "RATS" pop out of the phrase "Bureaucrats Decide." At the exposure speed of one thirtieth of a second, "rats" has likely not crossed the borderline of subjective perception and should not consciously register. The ad's creator said it was not his intention to create a subliminal ad, but rather to make the ad more visually interesting by flashing part of the word "bureaucrats" on the screen. "It was," he said, "just a coincidence" that the letters popping centerscreen out of "bureaucrats" spelled out the negative prime "rats." Such denials notwithstanding, Weinberger and Westen's (2008) experimental test shows an "affective contagion" effect such that on exposure to the subliminal "rats" prime candidates are evaluated negatively. In a follow up experiment, a photo of Bill Clinton primed evaluations of Governor Gray Davis in his 2003 recall election, with Republicans evaluating Davis more

negatively than Democrats. Both inside the lab and in the real world, unconscious priming effects like these are proving to be influential in how information is encoded, retrieved, interpreted, evaluated, and acted upon.

While the use of subliminal primes (Type 1 CUEs) in the laboratory provides the strongest experimental control and clearest demonstration of the automaticity of beliefs and attitudes and allows the researcher to rigorously test for the causal effects of unconscious events on both implicit and explicit attitudes and behavior, our endorsement of subliminal priming stops at the lab door, not on the airwaves or the campaign trail. Moreover, the use of truly subliminal priming in advertising is undoubtedly exceedingly rare. But the effects of consciously noticed but unappreciated (supraliminal) primes (Type 2 CUEs) are common throughout the social world and most obviously manipulated in the advertising realm.

The Stream of Political Information Processing

In the following chapters we set forth our affect-driven, dual-process model of the architecture and mechanisms that account for when, how, and why thoughts, feelings, and behavioral intentions come to mind automatically to promote the rationalization of political beliefs and attitudes. At this juncture let us outline our model in broad strokes, leaving for [Chapter 2](#) a detailed description of the architecture and processes that promote motivated reasoning. We take a constructionist approach whereby the content of one's thoughts and coloration of feelings change moment by moment in response to both noticed and unnoticed "priming" events that link changes in the immediate environment to changes in political beliefs, attitudes, and behaviors.

When an individual is exposed to a communication, the concepts in the message – whether consciously attended to or not – begin to activate the attendant concepts in *long-term memory*. Once a concept is activated, its activation spreads to all its related concepts (Collins and Loftus 1975), whether that connection is semantic or affective. As political communications generally involve a large number of concepts coming into perception in rapid succession (think of television ads combining still images, words, or video with a voice over narration, all of which would simultaneously activate associations in long-term memory), individual concepts become activated and reactivated in real time as they, and concepts related to them, are perceived. Then, in a matter of moments, the activation levels of current concepts and their associated concepts decrease to make ready for what information comes next.

At this point in the process, the second type of memory becomes relevant. In contrast with long-term memory, *working memory* has a severely limited capacity: only about seven concepts can coexist in working memory simultaneously (Barsalou, 1992; Rumelhart and Ortony, 1977; Simon, 1967). These concepts in working memory, in a very real sense, are what the individual is consciously thinking about at that time. Researchers have envisioned the

process of moving concepts from long-term memory to working memory through a pandemonium model (Larson, 1996; Neisser, 1967; Ratcliff and McKoon, 1996) in which activation is seen as a competition between all of the activated concepts, with those that are most activated, for whatever reason, being selected for further processing in working memory.

It is at this point that the parallel nature of the affective and semantic connections becomes critical. Those concepts that are most semantically implicated by the communication are of course likely to win the competition, and move into working memory. So, if an individual is reading a message about tax policy, the concept of taxes is going to be constantly activated and reactivated, as many of the concepts in the communication will either be about taxes directly, or about concepts closely related to taxes that will cause its further activation. However, the concepts related to taxes that are most likely to be brought into working memory, and therefore potentially enter the conscious awareness of the individual as relevant considerations, are those that are *both* semantically and affectively related to the concept. Suppose that taxes are viewed negatively, but there are an equal number of positively and negatively evaluated concepts that are semantically related to taxes (public works projects and tax refunds might be seen positively, while IRS audits and tax preparation might have a negative affective connection). Because the activation of the concept of taxes spreads both affectively and semantically, those concepts that are both semantically and affectively connected with the concept of taxes will most likely pop into working memory. So, when a message mentions taxes, a negatively viewed concept, the other associations that come into working memory are going to be biased in favor of other negatively viewed concepts. IRS audits rather than positively perceived public works projects are likely to win out.

Figure 1.4 presents an overview of our account of the stream of information processing from the initial unconscious registration of an event to the generation of an evaluative response. *The fundamental assumption driving our model is that both affective and cognitive reactions to external and internal events are triggered unconsciously, followed spontaneously by the spreading of activation through associative pathways which link thoughts to feelings, so that very early events, even those that remain invisible to conscious awareness, set the direction for all subsequent processing.* It is only at the tail end of this stream of processing that we become consciously aware of the associated thoughts and feelings generated moments earlier. It is at this moment that we experience what subjectively seems to be consciously initiated thinking and reasoning (Custers and Aarts, 2010; Libet, 1985).

Most of the key concepts and processes in our theory are represented in Figure 1.4, starting with the left to right causal directionality of processing through time. A stimulus event triggers the stream of processing, proceeding through affective and then cognitive mediators, and perhaps leading to the construction of evaluations of political objects and conscious deliberation. As a function of time, attention, and other factors, the likelihood of subjective

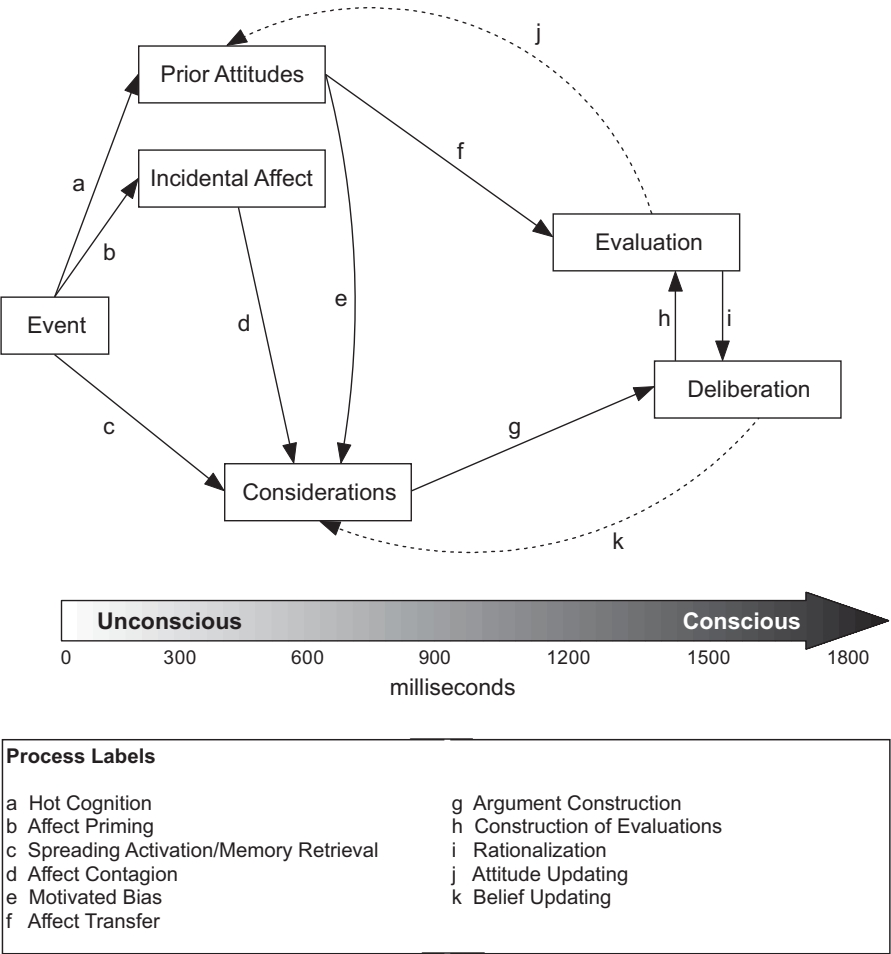


FIGURE 1.4. A Dual Process Model of Political Evaluation

awareness also increases left to right. Each arrow in the figure represents a theoretical process hypothesis. It is worth noting before we introduce these hypotheses that the conventional model of political reasoning involves only the c-g-h sequence in Figure 1.4: an event triggers the retrieval of cognitive considerations from memory, from which conscious deliberations are constructed, yielding reasoned evaluations.

While such controlled political cognition may sometimes occur, our dual process model claims that all thinking is suffused with feeling, and these feelings arise automatically within a few milliseconds (in our data as little as thirteen milliseconds) of exposure to a sociopolitical object or event. This is the *hot cognition* hypothesis that stands at the center of our theory of

motivated political reasoning. Affect is primary in our theory because it arises first in the stream of processing, is unintentional, and is difficult to control. Almost immediately, the decision stream becomes affectively charged, viscerally “hot,” and thereupon embodies our thoughts, providing proprioceptive feedback to mental processing (as shown, for example, by Damasio, 1994). Some of these feelings are attitudes that are intrinsic to the stimulus object (arrow a), while others are incidental or semantically unrelated to the stimulus (arrow b). Any subsequent considerations, deliberations, and evaluations are necessarily influenced by spontaneous affect. In terms of Figure 1.4, conventional political reasoning (causal path c-g-h) can occur only in the context of hot cognition.

Shortly after the arousal of positive and/or negative feelings, activation will spread along well-traveled associative pathways from, say, Obama to president to African-American to Democrat, thereby enriching our semantic understanding of the original stimulus. This is the *spreading activation* hypothesis (arrow c), well-established in cognitive psychology as the primary mechanism of memory retrieval. Note that many considerations may receive and send activation and thereby influence the stream of processing, but only a small number of highly activated considerations will reach conscious awareness – perhaps the 7 ± 2 chunks suggested in early psychological research (Miller, 1956).

In the context of just-aroused feelings, the retrieval of considerations will be biased in the direction of the valence of initial affect. This is the *affective contagion* hypothesis (arrow d) and the *motivated bias* hypothesis (arrow e). A flag, emotive music, an attractive candidate, or a celebrity spokesperson all influence the character of thought by favoring the retrieval of affectively congruent considerations while suppressing incongruent ones. Though it is possible for strongly associated concepts to reverse the direction of initial affect (as when initial positive affect triggered by a picture of John Edwards becomes strongly negative upon semantic recognition and retrieval of memories of his adulterous affair), it is more likely that initial feelings will “snowball” through the retrieval of increasingly congruent considerations, eventually driving deliberations and evaluations through indirect causal pathways. Spontaneous feelings can also cause evaluations directly through *affect transfer* (arrow f). A sunny day reliably drives more positive evaluations of life satisfaction (Schwartz and Clore, 1988). We have described how facial attractiveness directly drives positivity in addition to favoring the retrieval of more positive considerations. For evangelicals, Huckabee’s cross will promote a positive evaluation as well as prompting positive and more religious thoughts. The “rats ad” transferred negative affect directly onto evaluations of Al Gore. The twin influences of affect contagion and affect transfer are, we believe, among the most powerful and underappreciated sources of unexplained variation in studies of political evaluation.

With sufficient time and motivation, the retrieval of a set of considerations can trigger the construction of conscious deliberative reasoning given the motivation, opportunity, and cognitive wherewithal to query the immediate

affective response (Devine, 1989; Gawronski and Bodenhausen, 2007; Olson and Fazio, 2009). This process, labeled *argument construction* (arrow g) in Figure 1.4, will depend heavily on the earlier processes of hot cognition, spreading activation, and affect contagion. The central processes of motivated reasoning, including disconfirmation biases and the active counterarguing of counterattitudinal evidence, invoke these affective biases on memory retrieval (Taber, Cann, and Kucsova, 2009; Taber and Lodge, 2006). Conventional models of political thought view the conscious construction of arguments and reasoning as the foundations of public opinion and the guideposts to rational political behavior. We are skeptical.

Out of the grist of deliberation, citizens might *construct evaluations* (arrow h). That is, they might consciously build their evaluations of political figures, groups, or ideas from well-reasoned foundations, as in the conventional c-g-h model. In the context of hot cognition, affect contagion, and affect transfer, however, such cold evaluations will be exceedingly rare. The central place accorded to intentional rational evaluation in political science, a vestige of Enlightenment mythology in our view, continues to mislead our discipline, despite the valiant efforts of a few critics (David Sears and George Marcus come to mind).

Far more common, we believe, will be the reverse causal pathway from evaluation to deliberation. This *rationalization* hypothesis (arrow i) asserts that the causal pathways in Figure 1.4 that travel through unconscious affect, and in particular the affect-driven evaluation processes, cause most of our deliberation about politics. It is not our claim that citizens are incapable of rational thought in the traditional sense defined by links c-g-h. Evidence is accumulating, however, that attitudes and behavioral intentions – even behavior itself – arise from automatic, uncontrolled processes and are often set before we begin seriously “thinking” about them. This the case, deliberation serves to rationalize rather than cause.

The two dashed arrows in Figure 1.4 represent updating processes through which affect and considerations may be stored back to memory for future use. *Affect updating* (arrow j) allows the feelings and evaluations associated with current unconscious and conscious thought to be linked to objects in memory, where they can be the source of future hot cognition. For example, upon processing a newspaper story about Barack Obama’s handling of the BP Gulf oil spill, a citizen who was initially very positive about Obama may update her affect to be less positive or perhaps more ambivalent. *Belief updating* (arrow k) allows new beliefs or semantic associations to be stored in memory. This might include the creation of new memory objects (BP oil spill perhaps) or new linkages among objects (Obama and BP oil spill).

Notably absent from Figure 1.4 is any mention of emotions. In our theory, the appraisal of emotions follows and is directed by the arousal of valence affect and the motivating push of the concept’s somatic linkage. Appraised emotions (for a review of the appraisal literature, see Scherer, Shorr, and Johnstone,

2001) can be important mediators between aroused affect and subsequent processing, but for reasons detailed in Chapter 2 we will focus our attention on the causally prior processes of unconscious valence affect.

Most of this processing – the establishing of affect, meaning, and intentions – is subterranean, each process following one upon the other in about a second of time. An inkling of conscious awareness begins 300–400 milliseconds after stimulus exposure with a felt sense of positive and/or negative feeling, followed by a rudimentary semantic understanding of the concept, both of which are based entirely on prior unconscious processes. People can report simple like-dislike judgments in about 500–800 milliseconds and make simple semantic categorizations in 700–1,000 milliseconds, depending in part on whether the priming context for the categorization facilitates or inhibits comprehension. It takes somewhat longer (1,000–2,500 milliseconds) to provide a scaled response, and even longer to answer open-ended questions. Were we to ask a committed Republican to evaluate Secretary of State Clinton using a simple like/dislike button response, it would take about 700 milliseconds to press the dislike button. It would take significantly longer to report any cognitive associations to Hillary Clinton, that, for example, she is a woman, a Democrat, or mother. Affect precedes and contextualizes cognition.

Finally, given sufficient time and motivation, people may think self-consciously and reflectively about the object of evaluation and their own reactions. A point about conscious deliberation bears repeating: though deliberation will trigger new rounds of unconscious processing, it cannot go back and alter earlier processes and responses. In short, though we may feel we direct our thoughts and behaviors through conscious reasoning, deliberation is a product of unconsciously determined, affectively driven processes. Conscious deliberation and rumination is from this perspective the *rationalization* of multiple unconscious processes that recruit reasons to justify and explain beliefs, attitudes, and actions. It is possible, though difficult, to override implicit responses, as when we explicitly censor our socially unacceptable group stereotypes (Devine, 1989; Greenwald and Banaji, 1995), though it is not clear how fully we can control the “cognitive monster” of unconscious processing (Bargh, 1999). Our key argument and justification for the book title begins with, but then goes well beyond this primitive form of rationalization, to show how citizens’ snap judgments of likeability *as well as* their systematic thinking about political candidates and issues is motivated reasoning – a rationalization process driven by unconscious affective biases (for a parallel argument through the quite different lens of Affective Intelligence Theory, see Marcus, 2002). Emotions, like beliefs and attitudes, are reconstructed from what is made accessible to consciousness from unconscious memory processes, and in our model the positive and/or negative evaluative tally linked to an attitudinal object anchors the construction process.

For these and many more reasons, we are skeptical of the ability of citizens to reliably access or veridically report their beliefs and attitudes. Our

discipline's reliance on verbal self-report introduces a bushel basket of conceptual and measurement problems. In addition to well-known problems with the survey response (Tourengeau, Rips, and Rasinski, 2000), there is the obvious fact that the interview context, by design a sterile environment, is nothing like the immediate, situationally rich context that sparked the attitudinal response. In fact, it may well be the case that the simple act of asking questions promotes an intellectualization process that dampens the affective connection between thoughts and feelings (Epstein, 1972; 1992). These reasoned responses are no longer heartfelt, but affect negative *beliefs about the experience, not the experience itself*. Absent a somatosensory connection to the experience itself, the response is not embodied. Without a visceral boost the response is what Paula Niedenthal and her colleagues call a "cold, as-if emotional response" (Niedenthal, Halberstadt, and Setterlund, 1997; Niedenthal, Halberstadt, and Innes-Ker, 1999).

That the visceral experience need not be heart palpitating is demonstrated in a series of experiments carried out by Risen and Critcher (2011) testing a "visceral fit" hypothesis, the prediction that one's current bodily state – warmth, thirst, hunger – that "fits" the evaluation of a worldly event – here specifically aspects of global warming – will be judged more credible and likely. So, for example, feeling hungry will strengthen your estimate of the likelihood of famine, being thirsty makes droughts more probable.

In Study 1 on the Cornell campus during the months of September and October (with outside temperatures ranging from 49° F to 89° F) participants were taken outdoors for a psychophysical experiment ostensibly to measure the perceived height of various campus landmarks, then responded to a series of issue questions on eleven point scales, chief among them a CNN Poll question: "Which of the following statements comes closest to your view of global warming?" with the scale ranging from "Global warming is a proven fact" to "Global warming is a theory that has not yet been proven." Next, they reported their party ID and ideological self-placement (combined into a left-right index), and finally checked those terms they believed applied to their current physical state: hungry, thirsty, warm, tired, and chilly, while the experimenter measured the ambient outside temperature. Regressing belief in global warming on the outside temperature, left-right index, and the interaction term, ambient temperature proved to be as strong a predictor of belief in the validity of global warming, $\beta = .24$, $t(63)$, as ideology, $\beta = .22$, $t(63)$, and was not qualified by an interaction, with both liberals and conservatives reporting greater belief on warmer days.

In Study 2, to break the obvious diagnosticity of outdoor temperature to global warming, participants were randomly assigned to complete the survey in either a small heated room (81° F) or in an identical nonheated room (73° F). As in Study 1, both liberals and conservatives in the warmer environment were significantly more likely to believe that global warming was a proven fact, again without an interaction with ideology, although here ideology was a stronger

predictor than room temperature, ostensibly because the temperature indoors was not as readily associated with the outside weather. In other studies in this project pictures on the computer screen of desert scenes or snowy weather produced the expected viscera-fit effects. The favored explanation for the effects is that the bodily response makes it easier for people to imagine and simulate the belief. These “embodiment” effects are subtle and not readily recognized as influential and easily misattributed (Payne et al., 2005).

The Rationalizing Voter

Before turning to the empirics supporting this opinionation-as-rationalization argument, let us flesh out our line of reasoning for seeing citizens as rationalizing voters. Our model asserts that motivated reasoning – the systematic biasing of judgments in favor of automatically activated, affectively congruent beliefs and feelings – is built into the basic architecture and information processing mechanisms of the brain (Gazzaniga, 1992; 1998). Because both the spreading of semantic associations and biases favoring the retrieval of affectively congruent thoughts and feelings operate below awareness, the conscious, systematic construction of beliefs, attitudes, and intentions is necessarily dependent on those considerations and feelings that have been made available through unconscious processes. When called on to make an evaluation, state a preference, recount or justify an opinion, conscious introspection will not have access to the operative unconscious causal processes or many of the considerations that entered the decision stream unconsciously. Respondents, if pressed to account for their beliefs or attitudes, will as natural storytellers generate rationales that are more plausible than veridical (Clore and Isbell, 2001).

While the general principles guiding the role of accessibility and retrieval of information are well-known (Anderson, 1983), the implicit versus explicit distinction goes to the heart of our discipline’s problems in accounting for how, when, and why citizens think, reason, and act as they do. We expect that people will routinely rely on their spontaneously generated thoughts and feelings to explain their responses and behaviors, unless confronted by irrefutable evidence, social pressure, challenges to self-image, or interviewer pressure. And even here they will only experience these challenges as filtered through preconscious processes that have a built-in capacity for motivated bias. The experimental literature presents clear evidence that automatic processes underlie *all* conscious processing and are especially powerful determinants of top-of-the-head evaluations when

- affectively charged cognitions are available and strong;
- explicit measures are tainted by social desirability, deceit, or prejudice;
- one is under time pressure;
- attentional resources are otherwise engaged or distracted;

- an environmental event is noticed but not recognized as being influential; and
- one's behavior is not so consequential as to trigger such questions as "why did I think, feel, say, or do that?"

These situational and contextual factors appear to characterize the world of politics for many of us most of the time, where, typically, the consequences of our political beliefs and attitudes are distant and indirect, where uncertainty reigns, rumination is rarely called for, where one is easily distracted by rapid-fire TV images, and via selective media attention we infuse our thoughts with congenial cues.

Sometimes, of course, there is a feeling of unease with the considerations that come to mind, or a sensed dissociation of implicit from explicit thoughts, feelings, and intentions. If consciously conflicted, one may make the effort to resolve the conflict among and between thoughts and feelings (Gawronski and Bodenhausen, 2007). But there is now reason to believe that spontaneous activations are difficult to correct, even when people are encouraged to stop, think, deliberate, or actively try to work their way through a problem (Erisen, Lodge, and Taber, 2008; Forgas, 1995; Wilson, 2002). When constructing a response, the sample of retrieved considerations will likely be skewed in favor of affectively congruent associations. Because we are but dimly aware of the reasons for the thoughts that come to mind, those recollections entering the decision stream feel right, cannot be directly fathomed, do not typically produce a sense of dissonance, and consequently are not readily open to disconfirmation unless directly challenged.

While this argument of cognition as rationalization may seem radical, it is hardly new (see Achen and Bartels, 2006; Russell, 2003; Zajonc, 2000). Pioneering experimental work by Benjamin Libet (1985; 1993; 2004) demonstrates how consciousness lags behind even the intention to act. In a series of experiments, participants were asked to watch a sweeping clock hand, and report the moment when they made the decision to move a finger, while the researcher recorded their brain waves. Analysis of the EEGs revealed that a "readiness potential" to move the finger began approximately half a second before the conscious intention to move, but – and here is where the illusion of control comes in – the subjects retroactively predated their conscious experience by almost the exact amount of time it took the decision to reach consciousness (Libet, 2004), making the illusion of conscious control over these actions compelling. If the conscious decision to perform a physical action comes well after the intention has been formed, the notion that an individual's considered opinion precedes an automatic process is as likely an "illusion of conscious will" (Wegner, 2002).

These same processes apply to judgments. Zajonc (1980, 1984) found that even when people are able to give a reason for their judgments, the reasons they

give are often not the ones that informed the decision. This can be seen in the aforementioned “mere exposure” effect, in which subjects are found to prefer Chinese ideograms to which they had been previously exposed, without realizing that they had seen them before. Familiarity breeds liking. For our purposes, the most interesting aspect of the mere exposure effect is that, just as Libet’s subjects mistook when they had consciously initiated a simple physical motion so as to match it with the onset of unconscious initiation, Zajonc’s participants were able to give sensible reasons for liking one ideogram over another. Though they were consciously unaware of having seen some of the ideograms more frequently than others, they readily misattributed their preferences to the aesthetic value of the more frequently presented ideograms, rather than to the mere exposure effect where familiarity itself spurred liking. People are experts at rationalizing unconscious judgments. Moreover, even when explicitly told that they have been primed to evaluate the images in a pro or con way, people were still unable to overcome their automatic affective response (Winkielman, Zajonc, and Schwarz, 1997).

These effects – broadly speaking, the unconscious linking of feelings to thoughts to preferences to behavioral intentions – conspire to promote our view of the individual as more rationalizer than rational decision maker. Treating the citizen as a motivated reasoner will require a revolution in how we think about and model citizens’ mental representations of the world and the processes involved in the formation and expression of their political beliefs, attitudes, and behavior. When we limit ourselves to equating cognition with conscious awareness and the expression of preferences with the conscious integration of costs and benefits, as is the practice in political behavior research, it proves impossible to understand contemporary social, cognitive, and neuropsychology, and consequently makes it impossible to understand how, when, and why citizens think, reason, and act as they do.

At this juncture, we are highly skeptical of the ability of citizens to reliably and veridically access the sources of their beliefs, the reasons for their attitudes, their past, present, future intentions, and actions. Much if not most of our experience takes place outside our conscious awareness, and as our recollections fade from memory they are replaced by socially constructed rationalizations about how and why we as well as others think and behave. What recollections are activated depends on the set of preconditions operative in the environment *at the moment* and what’s going on inside the individual’s head *at the moment*. The key here is that once triggered, once the extant attitude enters the decision stream, thoughts are linked to feelings, feelings to intentions, and intentions to choices without necessarily triggering conscious or deliberative guidance.

Looking Ahead

Chapter 2 will detail our affect-driven, dual-process theory of motivated reasoning, and ensuing chapters will show how, when, and why the automatic

activation of affect spontaneously impacts the way citizens evaluate political leaders, groups, issues, and events. A basic finding, demonstrated in multiple experiments, is that feelings enter the evaluative process before cognitive considerations and immediately influence what thoughts and preferences will enter the decision stream. As we have already argued, this finding challenges the way we political scientists conventionally model the relationship between beliefs and attitudes – for most people most of the time the causal arrow flies spontaneously from affect to cognition, from preferences to thinking, from feeling to action.

As is common to the human condition, this “affect heuristic” is both a benefit and a problem, sometimes working well, at others leading us astray: on the plus side the primacy of affect promotes coherent thinking and attitudinally consistent behavior, but at one and the same time it is responsible for deep-rooted processes that bias how we think and reason. Where, when, how, and for whom conscious processing will successfully override the automatic intuitive response is the critical unanswered question that goes to the heart of all discussions of human rationality and the meaning of a responsible electorate. We leave a discussion of this paradox to the Conclusion but must forewarn the reader that we see no obvious resolution to the dilemma and cannot in good faith counsel as to when to follow the dictates of the heart.

The *John Q. Public* Model of Political Information Processing

In this chapter we set forth our theory of the architecture and mechanisms that determine when, how, and why unconscious thoughts, feelings, and goals come to mind to guide downstream political behavior. We take a constructionist approach whereby the content of one's thoughts, the coloration of feelings, the plausibility of goals, and the force of behavioral dispositions change moment-by-moment in response to "priming" events that spontaneously link changes in the environment to changes in beliefs, attitudes, and behavior. Far from the consciously directed decision-making assumed by conventional models, we see political behavior as the result of innumerable unnoticed forces, with conscious deliberation little more than a rationalization of the outputs of automatic affective and cognitive processing.

The Architecture of Memory

How we picture the world – our mental representation of self, other, and what is out there – is "the residue of a lifetime of observation, thought, and experience," both conscious and unconscious (Carlston, 2010: 38). A cornerstone of any model of political reasoning then is the citizen's preexisting knowledge and predilections. These long-term factors, functionally speaking, require a *long-term memory* (LTM) for storing facts, beliefs, images, feelings, habits, and behavioral predispositions, plus a mechanism for "moving" such conceptual objects as leaders, groups, events, and issues from LTM into *working memory* (WM) where they can be attended to (Barsalou, 1992; Rumelhart and Ortony, 1977; Sanford, 1987; Simon, 1969). Conscious attention is very limited, hence the need for heuristics, habits, and other simplifying mechanisms for thinking, reasoning, and doing (Cialdini, 2001; Lau and Redlawsk, 2006; Lupia, McCubbins, and Popkin, 2000; Kuklinski and Quirk, 2000). The important point is that those concepts and their connections processed in working

memory are strengthened and the resulting representation modifies the linkages in LTM.

The primary bottlenecks of consciousness, which stand in stark contrast to our much greater capacity for unconscious thought, are: (1) the small capacity of WM, which allows us to hold 7 ± 2 chunks of information in awareness at any one time (Miller, 1956); (2) the necessary displacement of old information in order to bring new information into conscious WM; and (3) strictly serial conscious processing, in which information must be processed sequentially (Payne, 1982). By contrast, LTM is vast and capable of highly parallel processing. The limits on conscious awareness are the primary reason for our being “bounded rationalists.”

Associative Memory. LTM is organized associatively, and it is useful to think of knowledge structures in LTM metaphorically as configurations of *nodes* linked one to another in a network of associations (Anderson, 1983; 1993), or if you prefer as neurons “bundled” together by weighted connections (Read and Miller, 1998; Smith, 1999). Were we able to tap into a citizen’s full political knowledge structure, there might be tens of thousands of conceptual objects (among them surely a node for Barack Obama), with a complex network of linked associations along well-trod pathways to the presidency, the Democratic Party, his characteristics and perceived traits, perhaps his stand on a few issues, and maybe an inferential abstraction or two, that, for example, he is somewhat liberal. Concepts are linked associatively to form beliefs (conceptual relations) and attitudes (affective relations), the strength of both varying from weak to strong. Moreover, memory objects vary in their *accessibility* – the ease with which a stored object lying dormant in LTM can be activated to influence information processing or even retrieved into conscious WM (Fazio, 2007).

Figure 2.1 sketches the architecture of a hypothetical citizen’s political knowledge structure, denoting different types of memory objects by shape, object accessibility by border thickness, and strength of association between nodes by the thickness of links. Here and throughout “object” refers to any sort of concept in LTM, be it a person, event, abstract idea, image, or your left foot. This particular example depicts knowledge about Barack Obama of a white American citizen, who identifies with the Republican Party. A variety of associations to Obama are shown, including perceived attributes or characteristics (ovals), groups or political persons (rectangles), emotions (rounded rectangles), and behavioral intentions (diamonds). Darker shaped borders signify more accessible objects. Links between conceptual objects represent beliefs, with darker links showing stronger beliefs; links to affective objects represent feelings or attitudes. All objects carry positive and/or negative affect, denoted by plus and minus signs respectively, and darker plus/minus signs represent stronger affect.

Attitude is a central concept, both historically and in contemporary research, of this and virtually all social-psychological models of human behavior, with attitude defined simply as the expression of one’s likes and dislikes, what one

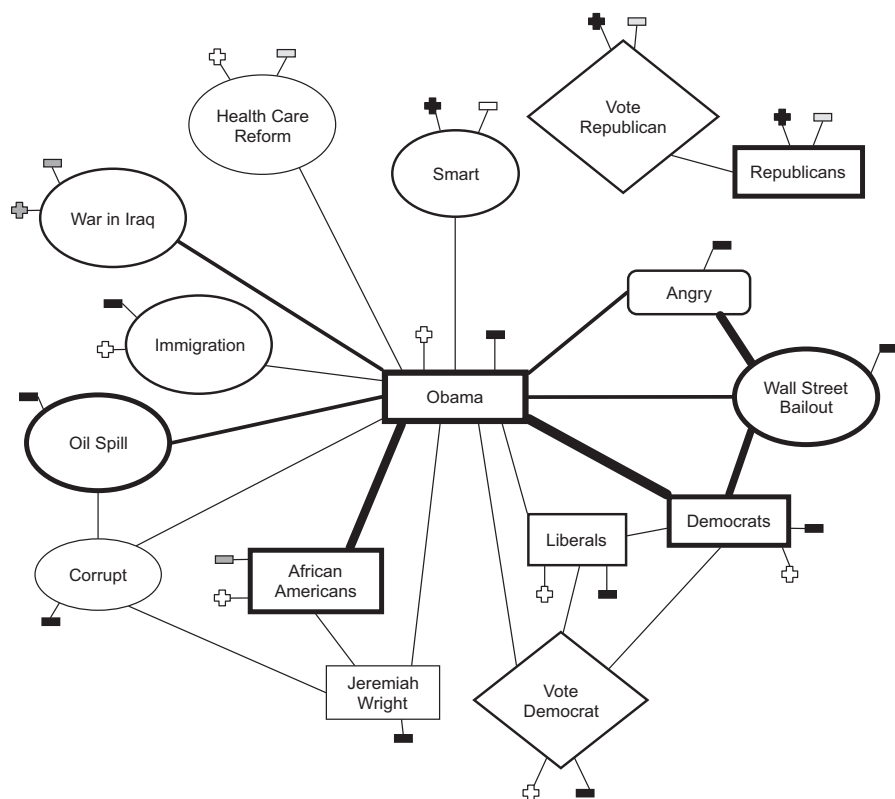


FIGURE 2.1. The Structure of Political Beliefs, Attitudes, and Intentions for a Hypothetical Citizen

favors or opposes, views positively or negatively (Petty and Briñol, 2010: 335). Here, following Fazio's lead (2007), we define attitude as an evaluative tally attached to an object in long-term memory. The hypothetical citizen in Figure 2.1 is generally negative or at best ambivalent toward President Obama. Obama is most strongly associated with Democrats and African Americans, and is appraised negatively for the oil spill recovery. This citizen is unlikely to vote for Obama, is especially angry (an appraised emotion that has been stored back to memory) for the Wall Street bailout, is ambivalent about the war in Iraq, and she has a weak voting preference for Republicans.

Our model, named *John Q. Public (JQP)* departs significantly from earlier node-link associative models (Anderson, 1983; Lodge and Stroh, 1993) in directly integrating positive and negative affect and goal-directed behavioral dispositions (in our example, the vote intention) into the model. From this Dewey-like "thinking-is-for-doing" vantage point, thought, feeling, and action are linked together, sometimes loosely, other times strongly, depending on one's history of experience.

Strong, affectively charged objects (in Figure 2.1, Obama, Democrats, African Americans, and to a lesser extent Wall Street) are routinely formed by evaluative (Pavlovian) conditioning where a previously neutral object is repeatedly paired with such affectively charged unconditioned stimuli as “evil,” “dangerous,” a frown, or a curse. Once formed, such evaluative associations can be activated spontaneously on mere exposure, without conscious consideration of their validity, and then prove to be remarkably resistant to countervailing information. The second type of relationship depicted in this knowledge structure are beliefs linking one concept to another, here for instance, that Democrats were responsible for the Wall Street bailout. Such propositions are more cognitively based and less resistant to disconfirmation, *if* the individual is aware of a contradiction, confronts strong countervailing evidence, is motivated to challenge the spontaneously activated positive or negative responses, and has the opportunity and cognitive wherewithal to adjust the attitude. As we will soon see, when concepts are affectively charged, as is the case for all self-related social concepts, attitude change is difficult at best.

It is worth emphasizing that implicit and explicit beliefs, attitudes, and goals are represented in much the same way in this memory architecture (Carlston, 2010; Gawronski and Bodenhausen, 2007; Ferguson and Porter, 2010). From this perspective – now the consensual view (Petty, Fazio, and Briñol, 2009; De Houwer and Moors, 2010) – explicit beliefs and attitudes require some level of deliberation and can be measured directly, typically by simply asking if the individual likes him/her/it. In contrast, people are either unaware of holding implicit beliefs or attitudes (Wilson, Lindsey and Schooler, 2000) or far more common they are unaware of the factor(s) that create, maintain, or change their beliefs or attitudes. Implicit and explicit beliefs, attitudes, or intentions are not different in how they are represented in long-term memory, but they do require different measurement strategies. Unlike the direct approach used for measuring explicit attitudes, measures of implicit attitudes or beliefs must be more indirect. Early approaches included the unobtrusive observation of bodily gestures, eye contact, and a variety of physiological responses (Webb et al., 1966). Now-a-days implicit beliefs or attitudes are routinely inferred from response latencies – the time it takes for a respondent to indicate a like or dislike, belief, preference, or intention (Fazio, 2007; Huckfeldt, Levine, Morgan, and Sprague, 1999; Lavine, Borgida, and Sullivan, 2000; Lodge, Taber, and Verhulst, 2011; Teige-Mocigemba, Klauer, and Sherman, 2010; Wentura and Degner, 2010).

Spreading Activation. But how is information moved from LTM into conscious WM? Spreading activation provides the mechanism. An object node in LTM switches from being dormant to a state of readiness with the potential to be moved into WM when it is activated, either by direct recognition or because it is linked to an associated object of thought. Figure 2.2 depicts the activation process, with the Y-axis representing the level of activation of a given node in LTM and the X-axis representing time in milliseconds. The rise time from

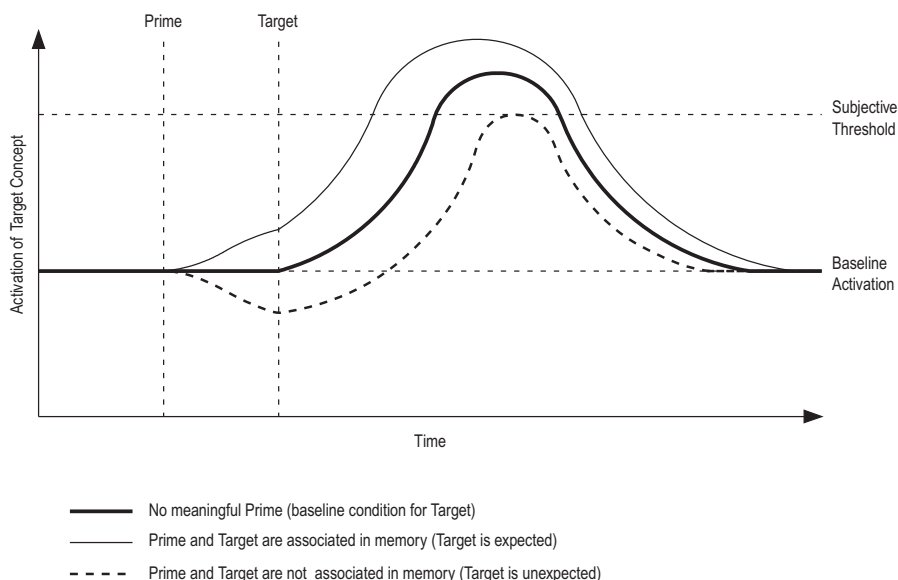


FIGURE 2.2. Activation of a Node in LTM

dormant-state to activation threshold is almost instantaneous (100–200 milliseconds). Were you thinking of George Washington’s false teeth? Are you now? Activation decays quite rapidly so that an energized node will drop back to its baseline level of potentiation in less than a second if there is no further source of activation.

In Figure 2.2, the distinction between unconscious and conscious processing corresponds to moving across the subjective threshold. All processes below that line involve LTM or other nonconscious processes, while processes above the line involve WM and conscious awareness. Imagine a person reading “President Obama” in a newspaper headline. Without perceptible effort, the concept Barack Obama is activated and energizes the network of links to related concepts (as seen earlier in Figure 2.1), where Obama primes our respondent’s strong semantic associations to Democrat and African American, as well as to her beliefs (he favors the war in Iraq), traits (he’s smart but corrupt), feelings (“he makes me mad”), and behavioral intentions (“I will vote against him”). For a few hundred milliseconds, these associated concepts remain in a heightened state of arousal.

It is useful and appropriate to think of priming through spreading activation as producing *preconscious* expectations. Figure 2.2 shows the activation of associations under different stimulus priming contexts. Consider again the activation of the concept Obama from a newspaper headline. Concepts associated with Obama in LTM receive spreading activation, thereby raising their potential so that any subsequent processing that passes activation to these

energized concepts may well drive them over the threshold into consciousness. If the citizen depicted in Figure 2.1 were primed by exposure to the concept Obama, it would facilitate (i.e., speed up) subsequent retrieval of African American or Wall Street Bailout (the “expected” activation curve in Figure 2.2). Priming concepts has several predictable effects on information processing: “expected” associations take substantially less processing time to activate and consequently will have a better chance of getting into WM, of being processed faster, and thereby of “framing” the perception, recognition, and interpretation of subsequent information down the processing stream.

Conversely, spreading activation can *inhibit* the processing of *unexpected* categories (the slowest activation curve in Figure 2.2). When a concept is encountered unexpectedly, more bottom-up processing is necessary before it may pass threshold and enter WM. If the word “Einstein” were heard immediately before reading a newspaper headline about Barack Obama, this would surely inhibit the recognition of such semantically unrelated concepts as Obama, Democrat, or oil spill, which would consequently take more time to recognize and more effort to process. Finally, the middle curve in Figure 2.2 represents a neutral or baseline case in which no “expectations” are created by a prime. The nonword letter string BBB, for example, conveys no semantic expectations, so would neither facilitate nor inhibit the recognition and categorization of subsequent concepts.

The long and the short of it: the closer the connections among and between beliefs, feelings, and intentions the more quickly coactivated associations pass threshold. This is the rationale for reaction time measures – the retrieval of related concepts are speeded up, distant concepts slowed down. Specific to attitudes, the activation of affectively congruent concepts is facilitated, incongruent pairings inhibited.

The strong implication of this architectural model is that all beliefs and attitudes will be constructed in real time from whatever cognitive and affective information is momentarily accessible from LTM. Keep in mind our earlier cautionary note against treating each conceptual node as a crystallized “point” in memory. Beliefs, attitudes, predilections, and intentions – all the stuff of consciousness – are constructed on the fly, in real time, at the moment of perception from whatever associations – whether conscious or not – make their way into Working Memory.

The best evidence for this construction process comes from neurological studies of patients with brain lesions showing that the naming, comprehension, evaluation, and functional use of concepts is distributed in different areas of the brain. So, listening to Mozart’s 40th *Symphony* activates hundreds of thousands of coordinated firings in the brain, with pitch processed in one set of neural regions, tempo in another, the timbre of a violin in another, all coming together seamlessly in a fraction of a second. Consider something more tangible – a tea cup. Visual agnosia is the inability of the brain to make sense of or make use of some part of an otherwise normal visual stimulus

and is typified by the inability to recognize familiar objects or faces. This is distinct from blindness, which is a lack of sensory input to the brain due to damage to the eye, optic nerve, or the primary visual cortex. Visual agnosia is often due to stroke affecting the posterior occipital and/or temporal lobe(s) in the brain. The specific dysfunctions vary depending on the type of agnosia. Some sufferers are unable to copy drawings but are able to manipulate objects with good dexterity. Commonly, patients can name the object, here a tea cup, categorize it, but cannot describe its function; or the reverse, be able to drink from it appropriately but not know its name or describe its uses. Lesion studies clearly demonstrate that even crystallized objects, your left foot, or here a tea cup, are not “things” in long-term memory but are concepts constructed from multiple brain modules at the moment of perception (Farah, 1999).

What is critical here is that neither implicit nor explicit beliefs or attitudes represent a solitary “point” in memory but are constructed in real time by the spreading of activation to cognitive, affective, and behavioral associations. People do not store in memory associative links for every conceivable category, subtype, or situation encountered in the past. The direction and strength of links to people, places, and things varies by one’s accumulated experiences across time and the immediate context. A puppy is generally not so “cute” when caught peeing on the rug. Intelligence is generally a positive trait but negative when seen in an enemy rather than a friend. If immediately prior to an assessment of George W. Bush a survey respondent were asked to evaluate the statement “All politicians are crooks,” the negative link to politician would predictably make the evaluation of Bush more negative.

Given that one cannot store attitudes in memory corresponding to every object in every conceivable context and because of the attitude construction processes for integrating affective associations, people reason and behave in ways that reflect the mechanics of spreading activation (Anderson, 1983; Boynton and Lodge, 1994; Gawronski and Sritharan, 2010). The effect of subtle contextual factors like question wording, question order, and interviewer effects on explicit beliefs and attitudes is well documented (Bishop, 2004; Tourangeau, Rips, and Rasinski, 2000) and is exactly what one should expect from this constructionist perspective where the momentary activation of cognitive, affective, and motivational associations will influence how objects and events are first implicitly and perhaps later explicitly perceived, conceived, and acted upon.

Seven Postulates Drive the Formation and Expression of Political Attitudes

Our theory can be captured by seven central claims: information processing is largely automatic, it is infused with feelings, it is embodied in physiological systems, it is impelled by affect, it is responsive to the environment through online updating processes, and it builds momentum through affect transfer and affective contagion.

Postulate 1, Automaticity. What people think, feel, say, and do is a direct function of the information that is momentarily accessible from memory – be it the recall of facts and feelings, a recollected experience, or the turning of goals into action (Greenwald and Banaji, 1995). Feelings of pride and in-group solidarity that swell when flags wave and patriotic music plays in the background of political events, the subtle confidence felt in the presence of tall political candidates or infatuation for attractive or charismatic ones, or the unease experienced by some voters at the prospect of an African American or female president can influence political thinking outside conscious awareness.

Research on automaticity demonstrates that beliefs, feelings, and behavioral intentions will, if “contiguously activated,” become so strongly connected in memory as to become *unitized* in a network of interdependent associations that enter the decision stream spontaneously on mere exposure to a “triggering event.” Automatic processes, in which thoughts, feelings, and intentions come to mind unconsciously, on a time scale of milliseconds, contrast with the more effortful processes people engage in when they have sufficient time, motivation, awareness, and the cognitive resources to deliberate. Process matters: with the repeated association of thought to feeling, beliefs become affectively charged; feelings motivate intentions, and plans direct behavior. From this perspective, Antonio Damasio (1999) is right in seeing the brain as a “thinking machine for feeling.”

Realizing that humans process information both consciously and unconsciously, theorists have proposed a conceptual distinction between attitudes that are the products of introspection and those that occur implicitly, outside of conscious appraisal. The labeling of one mode of processing as “conscious” emphasizes the reflective, deliberative character of responses to an “object” – whether person, place, event, thing, or idea – which generally (but not necessarily) involves verbal reasoning. Deliberative processes are cognitively effortful, demanding of attention, time consuming, and presumed to be based on an intentional search of memory for relevant facts and considerations. Such processes rely on the intensive use of WM, which as we have seen is severely limited in capacity and characterized by slow, one-chunk-at-a-time serial processing. Conversely, automatic processes – whether the immediate activation of cognitive associations (for example, Obama is a Democrat), the spontaneous activation of feelings (Republicans are evil; Democrats are dumb), or those habitual actions that operate “mindlessly” – are involuntary, fast, immediate, top of the head, and unlike conscious processes can be activated even when the individual’s conscious attention is focused elsewhere. These processes rely on the spreading of activation to associated cognitive, affective, and behavioral connections in LTM, which is vast and based on rapid parallel processing. People are frequently unaware of the specific situational and contextual factors that bring to mind the thoughts, feelings, and intentions that appear introspectively to be the outcome of a deliberative evaluation of the evidence (Wegner, 2002). Implicit processes, moreover, can, and oftentimes do, produce sound

decisions, sometimes better than those based on careful deliberation (Dijksterhuis and van Olden, 2006; Hofmann and Wilson, 2010; Verhulst, Lodge and Taber, 2010).

Triggering effects, whether consciously recognized or not, are ubiquitous in everyday life. Virtually all mental representations – be they words, pictures, sounds, or smells – appear to be “primeable,” that is, activated incidentally or unobtrusively in one context to influence one’s thoughts, feelings, goals, and even complex behaviors in another context, without the person necessarily being aware of having been influenced (Gawronski and Sritharan, 2010).

To call a process “automatic” it must satisfy four criteria (Bargh, 1997). It must be *spontaneous*, that is, the process or response must be triggered even if the individual is not consciously engaged in making an evaluation. The automatic influences on the judgment task must be *unconscious*. The response must be *uncontrollable*; once triggered, the process runs its course without conscious monitoring or guidance. And the process must *expend few cognitive resources*. In many familiar situations, as well as in such uncommon settings as a survey interview, automatic processes will directly impact the expression of evaluations, judgments, goals, decisions, and actions with little or no conscious or deliberative guidance. Given that implicit attitudes operate below conscious awareness, they cannot be measured directly, as typically by verbal self-report (De Houwer and Moors, 2010).

Unconscious priming effects have been demonstrated experimentally on virtually all higher mental processes:

- in making social judgments (Greenwald and Banaji, 1995);
- in attitude formation (Betsch, Plessner, Schwieren, and Guetig, 2001);
- the expression of beliefs (Neely, 1977);
- the expression of attitudes (Fazio, Sanbonmatsu, Powell, and Kardes, 1986);
- liberal – conservative ideology (Jost, Nosek, and Gosling, 2008);
- religious appeals (Albertson, 2011);
- trait inferencing (Newman and Uleman, 1989);
- self-esteem (Dijksterhuis, Albers, and Bongers, 2009);
- in-group and out-group identifications (Perdue, Dovidio, Gurtman, and Tyler, 1990);
- racial and gender stereotyping (Craemer, 2008; Devine, 1989; Dovidio, Evans, and Tyler, 1986);
- such national symbols (Butz, 2009) as the American (Ferguson and Hassin, 2007; Schatz and Lavine, 2007), confederate (Ehrlinger et al., 2011), and Israeli (Hassin et al., 2007) flags;
- such judicial symbols as a gavel, lady justice, and a judge’s robes (Gibson, Lodge, and Woodson, 2010);
- the making of moral judgments (Haidt, 2001);
- decision making across a variety of domains (Loewenstein and Lerner, 2003);

- the behavioral expression of egalitarian values in a competitive game (Bargh, Gollwitzer, Lee-Chai, Barndollar, and Troetsghel, 2001);
- corruption and the abuse of power (Chen, Lee-Chai and Bargh, 2001);
- evaluations of political candidates (Verhulst, Lodge, and Taber, 2010) and groups (Burdein, Lodge, and Taber, 2006);
- health decisions (Wiers, Houben, Roefs, de Jong, Hofmann, and Stacy, 2010);
- deliberate, reasoned belief systems such as political ideology (Jost, Nosek, and Gosling, 2008);
- deliberation about political policy issues (Erisen, Lodge, and Taber, 2007); and
- on a range of overt, goal-driven behaviors (Gollwitzer and Bargh, 1996), including consumer preferences and behavior (Perkins and Forehand, 2010).

Bargh (2007) draws an important distinction between *preconscious* and *postconscious* automaticity. In postconscious automaticity one is aware of the stimulus but not cognizant of its influence on thoughts, feelings, or behaviors, whereas in preconscious automaticity the stimulus is experienced below the threshold of conscious awareness so that the observer is not aware of having been exposed to the priming stimulus let alone able to appraise its costs and benefits.

Telltale evidence of postconscious automatic processing is routinely discerned in public opinion surveys, showing up most obviously as question-wording and question-order effects (Tourangeau, Rips, and Rasinski, 2000). For example, a *Washington Post* opinion poll asked a national sample of Americans in November of 2002, when President Bush's approval rating was in the mid-sixties, whether the country was headed "in the right direction" or "was seriously off in the wrong direction." Immediately before or after this question, they asked whether the respondent approved or disapproved of the job Bush was doing as President. A postconscious "Bush effect" is implied by the finding that 42 percent of those asked the Bush approval question first believed the country was headed in the right direction, whereas only 34 percent felt that way when the Bush question was asked second. Outside of conscious awareness, respondents' positive feelings toward Bush influenced their assessments of the state of the nation. One suspects that this postconscious "Bush effect" would be more horns than halo by mid-2004, when approval ratings had dropped twenty points.

Our theory predicts that priming effects – whether sparked by a President's name, upbeat music in the background of a commercial, the sound of prison doors slamming shut in the Willie Horton ad, or even having "rats" jump out of the word "bureaucRATS" – would produce the same biasing effect on information processing. As with flags and other symbols in the backdrop of presidential speeches, the more subtle and unobtrusive the "manipulation" the stronger the effect is expected to be as it would not trigger conscious reflection

(Gibson and Caldeira, 2009; Mendelberg, 2001; Schwarz and Bless, 1992; Schwarz and Clore, 1983).

Greenwald and Banaji's (1995) Implicit Association Test (IAT) is currently the most popular postconscious procedure for measuring automaticity (see demonstrations of the IAT at <http://implicit.harvard.edu>). A meta-analysis by Greenwald, Poehlman, Uhlmann, and Banaji (2009) compared implicit to explicit measures (184 independent samples, 14,900 experimental subjects), finding that correlations vary widely – from 0.18 to 0.68, with an average implicit-explicit correlation across attitudinal, judgmental, and behavioral measures of 0.27. When the implicit and explicit measures were collected in the same session (156 of the 184 samples), this average correlation rose to 0.36. These low-to-moderate correlations carry much the same power as do explicit-to-explicit measures on the same variables (Nosek and Smyth, 2007).

Latent-variable structural models on these data demonstrate that a two-factor model, with implicit and explicit attitudes as separate factors, is superior to a single-factor specification. “That is, despite sometimes strong relations between implicit and explicit attitude factors, collapsing their indicators into a single factor resulted in a relatively inferior model fit. We conclude that these implicit and explicit measures assess related but distinct attitudinal constructs” (Nosek and Smyth, 2007: 1). Support for the distinct contribution of implicit and explicit measures of political attitudes was found by Roccato and Zogmaister (2010) in their two-wave panel study of the 2004 Italian National Election. Employing both a comprehensive survey of explicit political attitudes and the Implicit Association Test, they found consistent relationships between both measures and vote intention, with the IAT in the first wave showing a significant, though modest, improvement in prediction to the actual vote.

The consensual view is that implicit and explicit attitudes are different processes working from a single underlying memory system (Gawronski, Strack, and Bodenhausen, 2009). It is not the case that implicit attitudes are stored in one way and explicit in another. The clear distinction is between conscious versus unconscious awareness. Explicit attitudes are consciously considered responses for which one has time to form a response. They will be influenced by lots of unnoticed factors, but there will be an opportunity for control and consciously reasoned thought. Implicit attitudes are affective responses to stimuli generated outside conscious awareness that one cannot control or consciously reason about. It is more likely that an implicit response reflects affect stored directly with a memory object (i.e., OL tag), but here too the response will be influenced by lots of extraneous factors. We think it is a mistake to think of one as more “true” than another. Although both are subject to certain types of bias, we will show that there is more opportunity for motivated bias when people explicitly consider and evaluate the considerations in mind, but this doesn't make explicit attitudes less “true” because motivated biases are likely to move in the direction of underlying affect.

The wide range of correlations between implicit and explicit attitude measures found by social psychologists is precisely what dual-process models predict, and what we expect given *JQP*'s architecture and built-in processes. That is, in some processing contexts or on some issues one may become consciously aware of the feelings and beliefs that drive a downstream behavior, while in other contexts these currents of thought will remain unavailable to explicit reporting. We will critique the extensive literature analyzing the relations between implicit and explicit measures of attitude at the end of this chapter, once the logic and measures of implicit processing have been set forth and in the conclusion following our empirical demonstrations.

More recent research goes beyond the automaticity of beliefs and attitudes to focus on the postconscious activation of complex social behaviors. In a now classic experiment, Bargh, Chen, and Burrows (1996) primed the concept "elderly" by having participants in the treatment group unscramble five word strings such as "bank sweater a knit she" and "lives Florida water in he" into grammatically correct four word sentences as quickly as possible. In addition to "knits" and "Florida," the stereotypic elderly primes (both positive and negative) were: worried, old, lonely, gray, selfishly, careful, sentimental, wise, stubborn, courteous, withdraw, forgetful, retired, wrinkle, rigid, traditional, bitter, obedient, conservative, dependent, ancient, helpless, gullible, cautious, alone, and of course, bingo. In the neutral treatment, the elderly prime words were replaced in scrambled sentences with words unrelated to the elderly stereotype (e.g., thirsty, clean, private). The behavioral dependent variable was the time, measured in seconds, it took subjects on completing the sentence unscrambling task to leave the lab and walk to the elevator. Those primed by the concept elderly took significantly longer than control subjects to walk the thirty meters to the elevator, even though none of the Type 2 primes for elderly referenced slowness of gait and the study participants were college students, not old folks. Yet, their mental representations of the elderly activated a rich behavioral script that included slow walking. Another study (among dozens of similar "ideomotor" demonstrations) primed one group of subjects to the concept "professor" and another to "soccer hooligan" and found (sigh of relief?) that the professor-primed group correctly answered more *Trivial Pursuit* questions than did those exposed to the hooligan primes (Dijksterhuis and van Knippenberg, 1998).

While much of the priming literature is focused on the stereotyping of others (in the United States, most commonly Jews before WWII, then African Americans, women, homosexuals, and more recently Muslims, the obese... and the list goes on [King, Shapiro, Hebl, Singletary, and Turner, 2006]), current research has taken an inward turn to examine the effects of one's own group identity on self-perceptions, expectations, and behavior. Here, for example, priming minorities' own group identifications has been demonstrated many times over to significantly lower their performance on standardized tests (Steel and Aronson, 1995). Called "stereotype threat," simply having African

Americans check off their race on a test form before taking a standardized test will result in a lower grade than if the identification question was asked after the test.

The same effect holds for women taking science tests. In a clever experiment demonstrating the subtle, insidious power of stereotype threat, Shih, Pittinsky, and Ambady (1999) worked with two stereotypes acknowledged commonly in the academy – Asians are good at math, and women are not. In their study, one group of Asian women completed a brief survey of attitudes toward coed dorms (which primed their gender identity), a second group was asked questions about their family history, language spoken at home, and such (to activate their ethnic identification), while the control group was asked neutral questions. Test performance on an objective math test matched the stereotype-threat expectation: performance was best in the Asian-identity condition, moderate in the control condition, and worst in the gender-identity condition. Here again, we see a postconscious effect: the women knew the cultural stereotype that was activated in the priming questions but were unaware of its potent effects on their performance.

In another series of remarkably subtle experiments suggesting the everyday importance of postconscious priming, Kay and colleagues (2004) investigated the effects of simple business Type 2 primes (for example, pictures of board-room tables, men and women's business suits, attaché cases) on competitive behavior. Their hypothesis: common objects carry implicit psychological meaning (i.e., business is competitive) that will prime the behavior of experimental subjects who are in the unobtrusive presence of these objects. The design across studies was to first engage participants in a postconscious business-related priming task, and then in an ostensibly unrelated second study engage participants in one or another behavioral task in which they could act cooperatively or competitively.

Study 1 asked treatment subjects to match business-related pictures to word labels, while control subjects performed the same priming task for such non-business objects as a kite, sheet music, and a toothbrush. All subjects were then asked to complete twenty-four word fragments, nine of which connoted competition, among them (w)in, (p)ower, wa(r), and one ambiguous fragment, “c_ _ p_ _ _ tive.” While none of the participants reported awareness of the relevance of the priming task to the word fragment task, those primed with business objects completed significantly more competitive word fragments than the control group. Moreover twenty-four of thirty-four treatment subjects saw “competitive” in the fragment c_ _ p_ _ _ tive, compared with just thirteen of thirty three in the control group, who were slightly more likely to see “cooperative.”

The next study looked beyond judgment to behavior. Following a similar postconscious picture-priming task, subjects were now asked to play an Ultimatum Game, in which one player chooses how to split \$10 with another

player in a one-time, take-it-or-leave-it proposition. Here too, the results show strong priming effects, even though the participants were consciously unaware of any connection between the picture primes and their subsequent behavior. All but one of the control subjects offered an even split, but seven of the eleven participants primed with business images offered significantly less.

The third study saves the best for last. The Ultimatum Game again, with all subjects in the role of choosing how much of \$10 to offer another (unseen) player, but now there is no picture-priming task. Instead, subjects write down their take-it-or-leave-it offer in one of two settings: half made their offer in a room with a long wooden conference table on which lay at the far end a leather briefcase and before them a black leather portfolio and wide-barrel, silver, executive style pen to write down their offer and then place the sheet in the briefcase. The other half of the participants performed the same Ultimatum Game task in the same room, but now a student's backpack replaced the briefcase at the far end of the table, a cardboard box substituted for the executive portfolio, and the take-it-or-leave-it bid was made using a wooden pencil. After making their offer, all were asked to list the factors that contributed to their offer. None indicated being influenced by any of the objects in the room, yet the results show significant effects of condition on offer, with all ten subjects in the backpack condition opting for a 50:50 split, while only six of twelve in the business setting did so. A significant priming effect was also found in the dollar amounts offered: on average, the paltry sum of \$3.89 was offered by those in the business setting, while a cooperative \$5 was offered by those in the scruffy student setting.

These studies and many more demonstrate the influence of unappreciated priming events on perceptions, social judgments, and behavior (Bargh, 2007). Of special note here is that in each case, study participants were consciously aware of the environmental primes, but were unaware of their biasing effects. Similar processes, we expect, permeate everyday life outside the laboratory (Bargh, 1997), especially so in commercial and campaign advertising, which are geared to promote positive messages without being heavy-handed (Singer, 2010).

In contrast, *preconscious* automatic responses – whether thoughts, feelings, motivations, or overt behaviors – occur spontaneously, within 300 to 500 milliseconds of a triggering event without conscious attention, awareness, intention, or monitoring. But who cares what happens in the blink of an eye? To answer this question, let us describe a trio of experiments by Dijksterhuis and Aarts (2003) that usher in themes that we will focus on when describing our own studies of automaticity in the evaluation of political leaders, groups, and issues. Dijksterhuis and Aarts set out to test the hypothesis, rooted in evolutionary theories of automatic vigilance, that people process negative stimuli more quickly than positive stimuli. Many studies have shown that negative events and objects demand more attention than do positives, but these Dutch studies

looked one step earlier in the process in asking whether negative stimuli are detected faster and easier at the *preconscious* level.

Study 1 tested whether participants would be able to detect positive or negative words flashed on a screen at the subliminal speed of 13 milliseconds, which is far too fast for conscious recognition. For half the trials, a positive or negative word appeared, for the other trials a nonword appeared. The subjects, who were fully informed in advance about the expected 50:50 frequency of words and nonwords but not of their valence, were asked after each trial whether they thought a word had or had not been presented. Not surprisingly at this subliminal exposure time, none of the subjects could consciously discriminate whether the stimulus flashed on the screen was a word or nonword, yet they correctly guessed significantly more of the negative than positive words. In short, what participants reported subjectively to be pure guesswork turned out to be systematically biased in favor of detecting negative stimuli.

But at what level did they perceive this negativity? Study 1 showed that negative words were detected faster, but it did not ask whether subjects pre-consciously recognized the valence of the words. Studies 2 and 3 take this next step, asking participants to press one key when guessing positive words and another for negative. Again, words were presented at thirteen milliseconds, but now either a positive or negative word was flashed on every trial, without nonword foils. Results confirmed expectations: the proportion of correctly identified negative words was significantly higher than correctly identified positive words, despite the fact that participants believed they were shooting in the dark. But can we yet be sure that this preconscious vigilance for negative stimuli was truly affective? Perhaps the semantic meaning of negative words is somehow processed faster than positive words.

Study 3 eliminated this possibility by asking participants to guess which of two same-valenced words presented explicitly on the screen was a synonym of the subliminally presented word. The results were striking: although subjects were significantly better able to detect negative than positive words, they were unable to reliably identify the synonym. The valence of concepts was identified pre-consciously but not the semantic meaning of the concepts. This then is a clear demonstration of preconscious affective processing and telltale evidence of a disjuncture between affective and semantic processing, with people able to “sense” that something is good or bad even though they are unable to tell you what it was they saw.

After five decades of well-replicated research, it is simply no longer tenable for those interested in understanding political attitudes, public opinion, campaigns, media, or vote decisions to ignore the effects of automaticity. Many, if not most, political scientists cling to an outmoded notion of rational behavior, in which citizens *cause* their issue stances, candidate preferences, and vote decisions through careful, intentional reasoning. Our research paints a very different portrait of the citizen as subject to the eddies and currents of innumerable priming events, some of which carry the potential to significantly alter

the course of information processing in ways that the citizen does not notice and cannot control.

Postulate 2, Hot Cognition. Conventionally, political scientists like their fellow social scientists have viewed the “holy trinity” of cognition, affect, and behavior as conceptually distinct and analytically separable, with cognition primary in causing both affect and behavior (Eagly and Chaiken, 1993; Kinder, 1998). Now a half century into the cognitive revolution (Eysenck and Keane, 1995; Lackman, Lackman, and Battlefield, 1979; Lindsay and Norman, 1977), we are finding it impossible to reliably tease apart thinking from feeling from behavioral intentions. Central to our affect-driven dual process model of attitude is the *hot cognition postulate* (Abelson, 1963), which brings feelings center stage in human information processing in claiming that all socio-political concepts are affect-laden (Bargh, 1997; Fazio et al., 1986; Sears, 2000). Fazio’s (1989) attitude theory, which is built into our theoretical architecture as represented in Figure 2.1, treats attitudes as object-evaluation associations stored in LTM memory, with their strength of association determining the likelihood that the evaluation will be activated on encountering the attitude object. Just as attitude objects can differ in the strength of their evaluative associations, people can differ in their chronic accessibility of evaluations (Lau, 1989). Specific to politics, all political leaders, groups, issues, symbols, and ideas thought about and evaluated in the past become affectively tagged – positively, negatively, or both – and with repeated coactivation an evaluative charge is linked directly to the concept in long-term memory. Affective tags represent the value of social objects as good, bad, or ambivalent.

With repeated evaluations an affective tag is linked to a concept and springs to mind spontaneously upon mere exposure to the associated object, thereby signaling the concept’s affective coloration. By election eve, most citizens will have formed impressions of the major candidates, parties, and issues and these feelings will be inescapably activated on their mere mention and will predictably come to mind most strongly and rapidly for those citizens who have given the most thought to the campaign. At the moment the president’s image on the TV screen passes threshold, one’s feelings about him come immediately to mind followed by his strongest cognitive associations. These accumulating positive and/or negative affective charges stimulate somatic changes in the body that will be experienced as positive or negative affect and then if strong enough to call for an answer to the question “why do I feel this way?” be labeled as a discrete emotion (for example, anger, fear, joy), with or without conscious awareness (Westen and Blogov, 2007).

The impact of context on evaluations follows directly from the *JQP* model. If “jobs” is primed for a working class citizen, “business” may be seen in a positive light, while in the context of “Wall Street,” “business” will likely be evaluated negatively. Note too that one is ambivalent when there are links to both positivity and negativity, as with “health care reform” and “the war in Iraq” in Figure 2.1. From our constructionist perspective, the evaluation of

an object represents the integration of multiple sources of affective information from the object itself (Obama is negative) as well as from its strongest associations (Democrats is negative).

The direct linking of feelings to concepts to goals and to behavioral intentions has profound implications for our conception of human information processing. The associative strength between an object (for example, politician) and its evaluation is conceived as varying along a continuum from nil – an object with little or no affective association, a “nonattitude” (Converse, 1964) – to a “crystallized attitude,” that is, an object with a strong, chronically accessible, univalent evaluation. Whereas weak and nonattitudes require effortful, piecemeal, bottom-up construction, the stronger the association between an object in memory and its affective tally the less time and effort needed to bring the attitude to mind, with objects carrying strong affective links activated automatically on mere exposure, without the observer necessarily being aware of even having perceived the triggering event (see Bargh, Chaiken, Gollwitzer, and Pratto, 1992).

Hot cognition helps solve the problem posed by the fourteenth-century French scholastic Jean Buridan, a student of William of Occam, who argued: “If a hungry ass were placed exactly between two hay-stacks in every way equal, it would starve to death, because there would be no motive why it should go to one rather than to the other” (quoted in Brewer, 1898, *Dictionary of Phrase and Fable*). Most humans, unlike Buridan’s ass, are equipped to solve such “equilibrium problems” by tagging the valence of goals which thereupon facilitates the making of quick, intuitive, directional choices (Wilson, 2002). Because affect permeates the entire decision-making system, beliefs, feelings, and actions will typically cohere (Thagard, 2000; 2006). When things go wrong, of course, there is a good chance that both thoughts and feelings will conspire to promote a misguided response.

This constructionist perspective implies that the evaluations a citizen might report in an opinion poll or vote choice reflect the integration of thoughts and feelings associated with one’s history of conscious and unconscious political evaluations. Immediately and without intentional control, a perceived candidate, issue, group, or idea is classified as either good or bad (Lodge and Taber, 2005; Morris, Squires, Taber, and Lodge, 2003), and in a matter of milliseconds, this evaluation facilitates a behavioral disposition toward the stimulus.

Note, again, that because unconscious processes are extraordinarily sensitive to contextual factors that easily escape conscious appraisal (Hofmann and Wilson, 2010; Niedenthal et al., 2005), the expression of beliefs and attitudes is context dependent (whether more or less so than conscious appraisals remains an unanswered question), and as will be demonstrated in upcoming chapters preconscious processes prove to be capable of integrating much more information into the decision stream than can be handled consciously. One consequence is that explicit attitudes appear to be (and are expected to be)

unstable over time and across situations because far more information has entered the decision stream than can be consciously processed.

Postulate 3, The Somatic Embodiment of Affect: In direct contrast to much of Western thought, which treats affect, feelings, and emotion as irrational intrusions that befuddle decision making, *JQP* follows the lead of recent neuro and social psychological evidence in connecting positive and negative feelings aroused by external events and internal thoughts to attitudes, goals, choices, and behavior. Hot cognition, the link from valence affect to cognition to preference to behavior, is viscerally monitored. Gut-level feelings automatically signal whether a person, situation, event, or option is seen as good or bad, threatening or rewarding. This embodiment of affect may be felt below conscious threshold as an intuition, or in other cases it may be experienced as intense arousal, demanding immediate cognitive appraisal for what and why I'm feeling this way. However it is experienced, whether consciously appraised or not, the immediate visceral response ensures that options will accompany perception, thereby facilitating approach or avoidance behaviors by signaling the prospect of pleasure or pain. Without the direct linking of feelings to thought to action, our beliefs and preferences would be cool, "as-if" experiences and as such likely to be weak, unstable, and poorly predictive of behavior (Niedenthal, Halberstadt, and Innes-Ker, 1999; Niedenthal, Halberstadt, and Setterlund, 1997).

Like so many advances on body-brain connections, contemporary research took its lead with patients who suffered damage to a particular area of the brain through accident, lesion, or stroke (Damasio, 1994; 1996; Damasio, Tranel, and Damasio, 1991). Those so afflicted provide us with analytic leverage for understanding what areas of the brain correspond to what functions: if you want to know what a particular area of the brain does, a time-honored starting point is to see what happens when it is damaged. In this case, the area of the brain of interest is the ventromedial prefrontal cortex (VMPC: in the middle of the brain on both sides, behind the eyes, right in front of the amygdalae). Individuals whose ventromedial cortexes are damaged retain their language and memory functions, all everyday cognitive abilities, but lose the ability to make use of their emotions to guide their social behaviors. They know what role emotions should play in such circumstances – how they and others should react – but simply don't feel the visceral tug pulling them in one direction or another. They live an "as-if" emotional life.

The VMPC is especially important because of its role associating knowledge about the environment with changes in bio-regulatory emotional states. In a social situation, the ventromedial prefrontal cortex associates similarities between the current situation and previous circumstances with the emotional context of the associations, creating what Antonio Damasio and his colleagues call a "body loop" (Bechara, Damasio, and Damasio, 2000). In essence, anticipatory visceral responses tell the individual that if she does *that*, she will likely feel *this*: if you criticize your partner's outfit, you'll regret it sooner, later, or

forever. This anticipatory affect is then used to guide behavior. So armed, the individual avoids behaviors that would have negative emotional consequences, and pursues those that have led to positive outcomes in the past. This process can be completely obscure to the individual experiencing it. In some cases, you may know and be able to articulate exactly why you like or dislike a particular person or choice, but often body-loop feedback simply takes the form of a good or bad gut feeling, the causes of which remain murky.

One of the earliest demonstrations of a disassociation of emotion from cognition dates to 1911 when the Swiss neurologist Claparede concealed a pin in his hand on greeting one of his amnesic patients with a handshake. The patient quickly withdrew her hand, but within minutes forgot the encounter. Shortly after, when Claparede reintroduced himself and offered his hand, the amnesic patient refused to shake, but when asked why could not remember being pricked with the pin. A contemporary demonstration of this disassociation of feeling from memory was carried out by Feinstein, Duff, and Tranel (2010), again with amnesic patients with damage to the hippocampus. Here patients, on watching sad and happy film clips showed the normal range of appropriate emotional expressions (smiling and laughing, frowning and crying), and continued to experience the emotion for as long as did the counter-balanced control participants, but they were unable just minutes later to recall much of anything about the film clips, and certainly not the reasons for their feelings.

A now-classic example of the disassociation of emotion from cognitive awareness comes from the Iowa Gambling Task (Bechara, Damasio, Damasio, and Anderson, 1994; Bechara, Damasio, Tranel, and Damasio, 2005). In this experimental procedure, participants choose 100 cards, one at a time (though they are not initially told how many cards they will draw), from four decks laid before them. On the face of each card, revealed only after it is chosen, is a monetary gain or loss, which is then added to or subtracted from the dollar amount accumulated. Two of the decks feature large payoffs, \$100 per card, but larger occasional losses of \$1,250. In the other two decks, the gains are smaller, only \$50 per card, but so are the losses: just \$250 per ten cards. On average, drawing ten cards from the two high-payoff, high-risk decks results in a net loss of \$250, while drawing from the low-risk, lower-payoff decks results in a net gain of \$250.

As would be expected, nearly all brain-intact participants begin by drawing cards in a relatively random order from the four decks. After about fifty draws, these players report having a hunch about which decks are better. After about eighty draws, they are typically able to articulate an understanding of the structure of the game and can explain their choice of cards in terms of a positive expectation about two of the decks. This is what we would expect from a deliberative solution to the game, and if this were the only mechanism through which players could intuit the game, they would after the eighty or so draws experience substantial losses. But this was clearly *not* the only mechanism at work, *nor even the one that actually drove behavior*, since these same players

avoided the risky decks long before they arrived at their first conscious glimmer of a hunch, beginning on average after just ten cards. Somehow, the Iowa gamblers (and participants in dozens of replications) were able to act on their real-time affective experiences in the game, without realizing they were doing so. How? The Iowa team suggested that physiological responses, monitored in the VMPC provide this mechanism, and they collected data to test their body-loop hypothesis.

While playing the gambling game, the Iowa team measured their participants' galvanic skin conductance, which tracks to the millisecond the participants' physiological stress levels outside of awareness and conscious control. In a striking confirmation of their hypothesis, after about ten cards stress levels rose sharply when participants considered the two high-risk decks. Most telling, this bodily response was precisely timed to their behavioral adaptations long *before* they could report a hunch about the risky decks. They were not consciously aware of what would likely be a good or bad outcome, yet these implicit impulses guided their behavior long before they became consciously aware of the relative costs and benefits of the card decks. At this early stage of the game, participants begin to play sensibly without even realizing that they're drawing from the good decks more than bad.

Individuals with damage to their VMPC simply don't get these somatic-induced hunches, and thus are unable to make use of the information about their visceral experience that others assemble unconsciously. Like the normal participants, those with bilateral damage to the VMPC begin by sampling a few cards from each deck, but unlike intact participants they don't gravitate towards the better decks as the game progresses. Instead, the longer the game continues, the more they draw from the high risk decks, which give them clearly observable higher payouts but large long-term losses. The critical difference seems to come from the anticipatory bodily reactions experienced by the normal participants: these somatic signals of stress do not occur among the impaired respondents (Damasio, 1996).

Findings like this are important to us not so much for what they tell us about the VMPC, but more for what they tell us about how people learn from their own visceral experiences: what is known as the *somatic marker hypothesis*. In essence, the brain uses feelings that have become associated with objects or behavioral options through good or bad past experiences as a visceral signal for the likely positive or negative consequences of an action. The intact participants in the Iowa Gambling experiments do not have a bad feeling about the high-risk decks because they calculated their expected values; rather they avoided these decks because they *felt* a stressful somatic response when they so much as moved a hand toward those decks. These intact participants decided "advantageously before knowing the advantageous strategy" (Bechara, Damasio, Tranel, and Damasio, 1997).

Our visceral embodiment postulate (with the somatic marker hypothesis a specific instance) is a 180 degree turnabout from the way most of us understand the relationship between consciousness and actions. Rather than

controlling our actions, consciousness often functions to interpret and rationalize the actions and processes that have already been carried out unconsciously. Nor is Damasio alone in putting forward such a model: Daniel Dennett (1991) and Douglas Hofstadter (2007) both arrived at similar conclusions from radically different starting points, as did Libet (1985, 2004) in his classic experiments involving conscious control of the timing of a simple motor behavior such as moving a finger. This sort of unconscious processing is troubling to many of us in political science and social psychology steeped in the belief that conscious thought precedes and dictates preference.

The visceral embodiment of affect allows the brain to use affect as real-time information to promote quick, efficient, spontaneous responses to what should be approached and what avoided. Over time, body-loop feedback helps to structure political knowledge, though by “structure” we have in mind associated connections among concepts rather than strict ideological constraint. It now appears that affect and cognition are *interdependent* systems, only separable in pathological cases. Among those of us with intact brains, all thinking, reasoning, and intentions, whether conscious or unconscious, are embodied by feelings of good or bad, like or dislike. Ironically, oftentimes tragically, the very same affective processes that contribute to rational action are also responsible for promoting bias in human thought, a theme we will pursue in following chapters.

Postulate 4, The Primacy of Affect. It is now well-documented that feelings enter the decision stream *before* any cognitive considerations come consciously to mind (Zajonc, 1980; 2000). Neurological studies of both mice and men suggest that the “affect system” follows a “quick and dirty” pathway that prepares organisms for approach-avoidance behavioral responses within 200–300 milliseconds of exposure and appear to enter the evaluation process spontaneously moments before cognitive considerations come to mind (Burdein, Lodge, and Taber, 2006; LeDoux, 1994; 1996; Morris, Squires, Taber, and Lodge, 2003). People feel their opposition to the Iraq and Afghan wars before any facts about the war (thousands dead and counting) come to mind, and these positive or negative feelings influence what cognitive considerations come consciously to mind. Even when one’s attention is focused elsewhere, automatic evaluative processes prepare the individual to make an affectively congruent response (Bargh and Chartrand, 1999).

The temporal primacy of affect over cognition seems perverse because it reverses causality in the conventional social science model and undermines the deliberative foundations of Enlightenment rationality. Feeling *before* cognition threatens normative and empirical standards for our understanding of political behavior because, if we are right, conscious deliberation is the wake behind the boat, while automatically stimulated affective and cognitive processes control the rudder. Rationale becomes rationalization.

As a preview of more to come, consider this clever demonstration of the direct link between feelings and approach-avoidance behavior: Chen and Bargh

(1999) instructed half their subjects to pull a joystick toward themselves when positive words appeared on the computer screen and push the lever away for negative words, while the other half received the opposite push-pull instructions. Results confirmed that subjects were faster to pull the joystick toward themselves for pleasant concepts and push away for unpleasant concepts, this a result we see as telltale evidence for the central role played by affect in triggering basic approach-avoidance behaviors.

The affective link to evaluations and choices helps prevent decision calculations from becoming so complex and cumbersome that choices would be impossible. Indeed, it is the primacy of affect that makes timely and effective decision making possible (Thagard, 2000). Feelings provide feedback about the unconscious processes that precede conscious consideration. Because automatic brain processing capacity is greater and faster than conscious appraisals, this “affect heuristic” (Slovic, 1999; Slovic et al., 2004; 2007), or better yet a “likeability heuristic” (Sniderman, Brody, and Tetlock, 1991) precedes and impels conscious processing. Without the weighting of goals by feelings no option would be more important than another and we would consequently end up like Buridan’s ass, unable to choose among preferences. But how are these feelings updated and in response to what?

Postulate 5, Online Updating of Evaluations. Benjamin Franklin (1779) proposed a classic decision-making strategy when advising his grand nephew on choosing a marriage partner:

Follow your own judgment. If you doubt, set down all the reasons, pro and con, in opposite columns on a sheet of paper, and when you have considered them two or three days, perform an operation similar to that in some questions of algebra; observe what reasons or motives in each column are equal in weight, one to one, one to two, two to three, or the like, and when you have struck out from both sides all the equalities, you will see in which column remains the balance. It is for want of having all of the motives for and against an important action present in or before the mind at the same time, that people hesitate and change their determinations backwards and forwards day after day, as different sets of reasons are recollected or forgot, and if they conclude and act upon the last set, it is perhaps not because those were the best, but because they happen to be present in the mind, and the better absent. This kind of Moral Algebra I have often practiced in important and dubious concerns, and tho’ it cannot be mathematically exact, I have found it to be extremely useful. I am your ever affectionate Uncle.

Note the commonsense wisdom of “Franklin’s Rule.” Making a list of pros and cons and checking them twice, he would probably expand the number of considerations entering the mix, perhaps stop himself from jumping to conclusions, and conceivably (although we think not) weigh and integrate the most important considerations most heavily into the equation. Kelley and Mirer (1974: 574), mirroring Franklin’s Rule, state the case for such *memory-based* models of vote choice: “The voter canvasses his likes and dislikes of the leading candidates and major parties involved in an election. Weighing each like and

dislike equally, he votes for the candidate toward whom he has the greatest number of net favorable attitudes. . . . ”

Nearly all theories and empirical studies of public opinion assume memory based processing. While they differ in what memory “considerations” come to mind – whether a candidate’s partisan affiliation, endorsements, personality traits, or policy preferences – they all suppose that a citizen’s evaluation is a straightforward function of consciously retrieved thoughts. Were this generic account of the evaluation process a faithful representation of how citizens actually construct their preferences, it would indeed make the pollster’s life simpler. If citizens could provide a more or less veridical account of what campaign events led them to favor one candidate over another, we could simply ask respondents for their likes, dislikes, and reasons for preferring one candidate over another and not bother examining the on-going psychological processes that convert campaign events into political preferences and vote choice. All we would need to do is have respondents recount the considerations that come to mind, which is exactly what is done in the National Election Studies.

To go beyond the tautological “voters vote for the candidate they like best,” the analyst must assume that the mix of recollections evoked by the open-ended, like-dislike, and issue-proximity responses stand as valid expressions of beliefs, attitudes, and reasons for vote intentions. True enough, across hundreds of survey-based studies, we do find a positive correlation between the self-report measures and candidate preferences. Respondents regularly vote for the candidate they tell us they like best and/or is closest to them on one or a few issues. If the criterion for success is prediction (or more commonly in practice, postdiction), all is well and good. If, however, the aim is to work our way out of the black box, to learn when and how citizens go about forming and updating their impressions of candidates and issues, the first question to be asked is how much credence can be placed in the citizen’s recall of likes and dislikes or placements of self and candidates on issue proximity scales as bona fide descriptions of how preferences and choices are formed and updated.

A major difficulty with memory-based models based on survey responses is that they are unable to distinguish among various psychological mechanisms underlying the judgment process and consequently cannot tease out the causal ordering of effects. There’s an oddity in the evidence that alerts us to the problem: in contrast to the survey results, experiments in social and cognitive psychology routinely report weak correlations between the mix of pro and con evidence recalled from memory and the direction and strength of various social evaluations (Anderson and Hubert, 1963; Hastie and Park, 1986; Lichtenstein and Srull, 1987; Lodge and Stroh, 1993; Lodge, Steenbergen, and Brau, 1995). When the researcher knows exactly what information respondents actually see, as is the case in these experiments, we find that the considerations people remember about other people, places, and events and the reasons they give for their preferences provide a poor fit to their evaluations. People tend to recall their good-bad, like-dislike global assessments, *not* the specific considerations

that actually went into their evaluations. At the very best, the recollections represent a biased sampling of the actual causal determinants (McGraw, Lodge, and Stroh, 1990; Reyes, Thompson, and Bower, 1980). At worse, recollections may simply reflect rationalizations dredged up to support the global judgment constructed earlier in the information processing stream (Pratkanis, 1989). In both cases the correlation between memory and judgment is spurious, and causal “explanations” based on these explicit “recollections” have the arrow going the wrong direction.

There is now a great deal of evidence showing that people form impressions of others automatically, anchor on this early impression, and adjust insufficiently to later information (Uleman and Bargh, 1989). By contrast, the recall of considerations is most strongly influenced by the most recently processed information. Information that comes early in a political campaign will more strongly influence the reception, interpretation, and evaluation of new campaign information, even though later information is typically better remembered (Lodge, McGraw and Stroh, 1989; Zaller and Feldman, 1992). This pattern of primacy effects on spontaneous impression formation but recency effects on explicit memory renders recall-based measures suspect as indicators of why people favor one person or idea over another.

The failure to find empirical support for the memory-causes-judgment hypothesis across a broad range of topics and tasks under experimental conditions where the researcher has control over the content of the message has sparked interest in what is called *online processing*. Online (OL) models hold that beliefs and attitudes are constructed in real time, at the moment of comprehension, when an object is before your eyes, so to speak (Anderson, 1965; Hastie and Park, 1986; Lodge, Steenbergen, and Brau, 1995). When people form or revise their overall impressions of persons, places, events, or issues, they are found to spontaneously extract the affective value of the message, and then within milliseconds integrate their appraisal of the object into their prior evaluation, all without any conscious query of memory for a set of considerations on which to compute an updated evaluation, as prescribed in Franklin’s Fable. This “running” OL tally, representing an automatic integration of all prior evaluations of the object, is then restored to long-term memory where it is readily available for subsequent evaluations (Cassinio and Lodge, 2007). From this OL constructionist perspective, affect infuses the encoding, retrieval, and comprehension of information, its expression as a preference, and readies us to act aversively or appetitively in accord with our feelings (Ito and Cacioppo, 1999). The OL tally is an elemental processing heuristic, more “primitive” we think than other such heuristics as partisan identification and the stereotyping of others known to guide impressions.

Because the online updating of attitudes necessarily moves forward in time, with existing feelings ever-present to influence subsequent processing, early information will have a greater impact on attitudes than later information. This powerful effect of primacy on impressions was first suggested by Solomon Asch

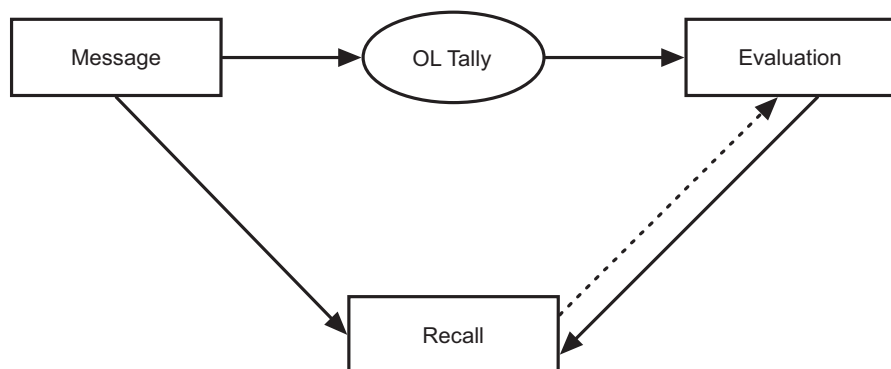


FIGURE 2.3. Online Model of Evaluation Processes

(1946), who argued that the very same descriptors of a person would produce a different holistic impression, depending on their order of presentation. Try the following thought “experiment” for yourself: On a scale that ranges from “highly favorable” (+4) to “highly unfavorable” (−4) evaluate person A, who is described as:

Faultfinding, awkward, cool, sentimental, athletic, and smart.

Now, to clear working memory count backwards by 7s from 100 until you reach 65, then evaluate person B, who is described as:

Smart, athletic sentimental, cool, awkward, and faultfinding.

Experimental participants evaluated person A as “slightly unfavorable” (−0.7) and person B as “moderately favorable” (+1.4), despite the fact that only the order of the trait descriptors changed (Anderson and Banios, 1961). First impressions count.

From this online perspective, as depicted in Figure 2.3 (adapted from Lodge, Steenbergen, and Brau, 1995), the appraisal of campaign messages is directly integrated into a summary evaluation (OL Tally), which directly informs the candidate evaluation. The recall of campaign messages contributes only weakly to the summary evaluation (dashed line) and is likely to reflect tally-driven rationalizations of the OL impression (the solid arrow from evaluation to recall).

Summary impressions bias recall in two distinct ways, which we will document in ensuing chapters. First, as noted earlier, an affective tally enters the processing stream earlier than do a concept’s semantic associations (you feel you like or dislike Barack Obama before you remember he is a Democrat), and consequently affect anchors judgments, which trigger why-do-I-feel-this-way rationalizations. This is the indirect path from the OL tally through the evaluation to recall in Figure 2.3. A second, more direct influence on memory is the “affective contagion effect” in which one’s OL tally biases the sampling of

recalled information in tally-consistent ways (Erisen, Lodge, and Taber, 2008; Rahn, Krosnik, and Breuning, 1994). This would be a direct arrow from the OL tally to recall in Figure 2.3. Moreover, semantic information is subject to an exponential forgetting curve (Ebbinghaus, 1885; 1913), while evaluations of an object are relatively stable over longer periods of time. In short, affective associations tend to persist in long-term memory and over time will outlive the semantic associations to an object leading to an ever-increasing infusion of affect.

To make their central point, early descriptions of online information processing drew too sharp a dichotomy between memory-based and online processing (e.g., Lodge, McGraw, and Stroh, 1989; Lodge, Steenbergen, and Brau, 1995). An either-or view is theoretically flawed and empirically untenable (Kim, Taber, and Lodge, 2010; Lavine, 2002; Redlawsk, 2001; Taber, 2003). The confusion stems from the failure to clearly discriminate encoding from retrieval effects. Recall the semantic network model in Figure 2.1, where affect is linked directly to concepts in LTM. Affective tags are attached when an object (person, group, place, event, issue, or abstract concept) is first evaluated and strengthened with each replication through the automatic online process we have just described. After but one or two evaluations, a concept is “hot,” affective charged (Lodge and Taber, 2005). From this point onward, affect and cognition are unitized in memory and difficult (we believe impossible) to disentangle in practice, though they remain conceptually distinct.

In a survey context the measurement of OL processing of candidate evaluations proceeds in stages. First, participants are asked to evaluate all the information that will subsequently be presented in a candidate message, plus many other pieces of information that will not appear in the message so as to later check for rationalization effects in recall. Next, there is a distracter task, perhaps questions asking for demographics, to thwart short-term rote memory of the items and their evaluations. Then, participants read about one or more candidates or issues, typically embedded in narrative form as a newspaper article or newscast. This is the candidate message. Fourth, participants are asked to evaluate the candidates. And finally, after another distracter or better yet a longer delay, participants are asked to recall the information in the article, followed by probes of recognition-memory asking for details about the candidate’s demographics, issue positions, and any trait inferences they may have inferred. Compute the correlation between an integration of the likes/dislikes for *all* the information presented in the candidate message and the reported candidate evaluation. Compute the correlation between an integration of the likes/dislikes for the *recalled* information and the reported candidate evaluation. To the degree that the first correlation is stronger than the second (or that the first statistical relationship survives inclusion of the second in a multivariate model), there is evidence of online processing.

Note that within this survey context the measurement of the OL evaluation is explicit; participants are asked directly for their preferences. But there is now

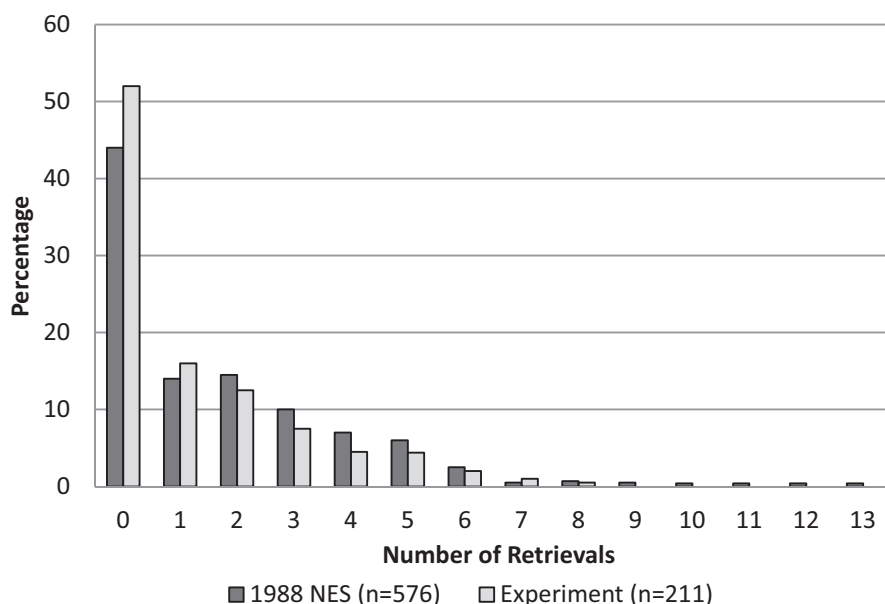


FIGURE 2.4. Memory Retrieval in Two-Candidate Election

empirical evidence and theoretical rationale for believing that OL processing is automatic, that people evaluate, update, and integrate their evaluations into a summary judgment effortlessly, outside of conscious awareness.

A well-replicated finding (Bassili, 1989; Hastie and Park, 1986; Lodge, Steenbergen, and Brau, 1995) is that whereas people can integrate lots of complex information in real time, even without being explicitly instructed to evaluate the object (e.g., candidate), they prove unable to recall much of this information after a short distracter task. And a day or two later the number and accuracy of recall no longer predicts either the information in the message or the summary evaluation. When Lodge, Steenbergen, and Brau (1995) looked at how much of the presented campaign messages their experimental participants could recall, the results were consistent with the findings from public opinion surveys. As shown in Figure 2.4, citizens forget... a lot, with about 54 percent of our respondents unable to recollect a single issue that either of the two candidates had addressed. The modal number of recalls for the policies' gist meanings (for example, "Candidate Williams opposes abortion") was zero, while recall for his more complex issue position ("... except in the case of rape or incest") was worse still, with more than 75 percent of the subjects unable to recall even one qualifier for either of the two candidates.

This dismal level of recall of campaign information by our subjects is not different from what researchers typically find in surveys about real-life candidates (Delli Carpini and Keeter 1991; Erskine, 1963; Neuman 1986; Smith,

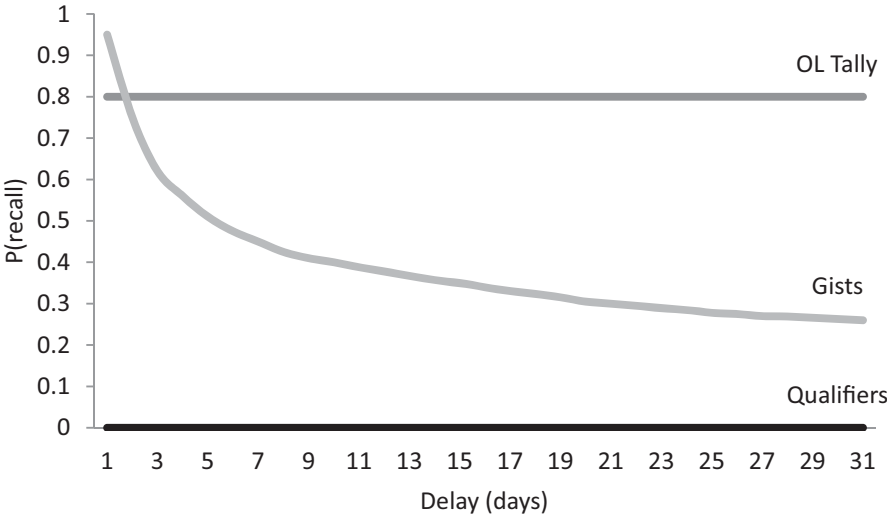


FIGURE 2.5. Online and Memory-Based Forgetting Curves for Democratic Candidate

1989). For instance, as shown in Figure 2.4, when we contrasted the number of recalls in our 1995 experiment with those for congressional candidates in the 1988 NES, we found a striking congruence. The loss-curves are virtually identical. In both samples the modal number of recalls is zero, with very few people providing more than two recollections of the campaign. Voters appear to forget much of the campaign information to which they were exposed, even after year-long campaigns.

Recall how severely limited conscious processing is compared to the much greater capacity of unconscious processes. Over the last decade or so experimental evidence has accumulated to suggest, if not yet demonstrate to everyone's satisfaction, that unconscious processing allows people to take in far more social information than can be processed consciously. The experiment by Lodge, Steenbergen, and Brau (1995) included a variety of manipulations to check for what goes into the OL Tally and so provides evidence as to the relative capacities of unconscious and conscious processing. To examine the effects of recall, all subjects were re-interviewed, this time by phone, one to thirty-one days after the experiment. Two striking results stand out in Figure 2.5: memory of the affective tally persisted over the thirty-one-day period, while recall of the candidates' gist policy proposals decayed exponentially and any specific policy details were essentially zero from day one.

Now to the key point from the Lodge, Steenbergen, and Brau experiment: in every case subjects' average like/dislike ratings of the candidates' seven policy statements in the brochure (i.e., their OL tally) strongly predicted their candidate evaluations, while the information they could recall did not predict their candidate evaluations, even for those who were instructed to think hard about

the candidates while they read the brochures. What is even more problematic for memory-based processing, the considerations subjects voiced when asked for their likes, dislikes, and reasons for their preferences are strongly biased by their summary attitudes, and many of their “memories” prove to be little more than rationalizations and projections stemming from their OL-based tally favoring one candidate over the other. What we see here is evidence that people are able to integrate far more information into preferences than they can recall and more importantly more complex information than presumably they can process consciously.

A compelling test of postconscious automaticity in online processing was carried out by a team of German social psychologists (Betsch et al., 2002) who had their subjects watch a series of thirty videotaped TV commercials, which they were told they would have to later recall and evaluate. Simultaneously, these subjects performed a second, cognitively demanding “distractor” task: they were asked to read aloud the changing stock prices of five hypothetical companies presented on a crawler at the bottom of the TV screen. Though participants were led to believe that their ability to remember and evaluate the TV commercials under pressure of an irrelevant distraction was the primary purpose of the study (their recall of the commercials proved to be very good), the researchers were actually interested in how the viewers would track the stock ticker outside their focal attention. In a surprise test, subjects were asked their preferences for the five companies. As predicted by the online processing model, participants were unable to recall much if any of the pertinent stock information, yet their summary, rank-ordered evaluations correlated positively and strongly with the actual performance of the 5 companies as had been reported in the stock ticker prices they could not remember. These results point to the spontaneity of online evaluations: subjects accurately evaluated the companies’ stock performance even when their attention was actively focused on an unrelated, attention-demanding task.

In sum, the evidence is now clear that people automatically update their attitudes toward a variety of social and political objects at the time they encounter relevant information. The process is autoregressive: the summary judgment is not based on a fresh look at all the evidence, but rather on the OL tally and the biased assessment of messages at the instant they are received (Huckfeldt et al., 2011). This updating process does not require conscious intervention and appears to engage our substantial capacity for the unconscious processing of large amounts of information. As such, citizens may be more responsive to the complex, high-volume information available in the political environment than conventional memory-based models suggest. This greater responsiveness, however, may come at the price of motivated bias in perceptions and judgments.

Postulate 6, Affect Transfer. The direct linking of positive and negative feelings to objects in memory is often the result of *affect transfer*, in which current affective states become associated with currently activated objects. The simple pairing of positive or negative cues with an object will tend to transfer

affect to the object in a form of classical conditioning that operates below conscious awareness. The sunny day effect is an oft-noted case in point. Life and love seem sweeter on a beautiful day, and even mundane objects may gain cachet. Advertisers routinely exploit affect transfer to manipulate feelings about candidates or commodities. Positive cues work by making the perceived object more likeable, while negative responses follow quickly upon exposure to negative cues. The effects are simple, direct, immediate, and spontaneous. Other responses, say an emotional appraisal, may or may not follow upon the immediate affective response.

The sunny day effect is a good example of *extrinsic* affect transfer in which feelings generated by an event *unrelated* to the object are transferred to the object. Common also is *intrinsic* affect transfer, in which positive or negative feelings that are relevant to an object become associated with it through simple pairing. For example, when a friend takes an issue position or wears a lapel pin supporting a political candidate, one's affect toward that position or candidate may become directly transferred to the friend (or vice versa). Similarly, when a presidential candidate announces his running mate he is likely hoping for some affect transfer from the bottom to the top of the ticket. Note that because affect transfer occurs on mere association of two objects or events, it can move in either direction, though a reasonable prediction would be that affect would more likely transfer from a strongly to a weakly felt object. The distinction between extrinsic and intrinsic affect transfer is in many ways a normative one, because the relevance of objects or events to each other is somewhat subjective, while the underlying simple conditioning process is the same.

It is important to recognize that our theory goes beyond a simple good-bad bipolar response in allowing for people to hold *both* positive and negative affect toward a political leader, group, or issue. The theoretical rationale here is the independence of positive and negative affect (Cacioppo, Gardner, and Berntson, 1997). One's evaluation can be ambivalent, with both positive and negative reactions activated, if not simultaneously, then triggered within milliseconds of one another. As will be demonstrated repeatedly in the upcoming experiments, this ambivalence has important consequences for how objects are evaluated and how information is processed.

A head's up is appropriate here: several scholars of the affective intelligence school (Brader, 2005; Marcus, Neuman, and MacKuen, 2000) argue against the affect-transfer hypothesis in claiming that positive and negative emotions, in particular enthusiasm versus anxiety, promote different behavioral responses above and beyond their valence. Perhaps. But our claim is that affect enters the decision stream first and subsequent emotional appraisals are heavily anchored by the initial valence response; such appraisals are often part of a postbehavioral rationalization process rather than the impetus for said behavior. Moreover, the empirical evidence provided in support of the causal importance of enthusiasm and anxiety may be better explained by affect transfer. When valence is controlled for in these analyses there is little left-over variance to be explained

by like-valenced discrete emotions. We will return to this debate in the conclusion after showing multiple experimental demonstrations of the direct effects of positive and negative affect on the evaluations of political leaders, groups, and issues, and on the in-depth processing of campaign information.

Postulate 7, Affect Contagion. In our theory, the affective tallies tagged to social concepts and updated through online processes inevitably color all phases of the evaluation process, sometimes explicitly, other times not, sometimes for good, other times not. Hot cognition and affect transfer provide two direct mechanisms. If these transitory effects were the full extent of the influence of affective online tallies on evaluations they would still be important if that momentarily activated affect is integrated in real time into evaluations and thereafter anchors subsequent judgments. However, hot cognition and affect transfer are not the only ways that feelings alter thought processes. *Affect contagion* will also influence political reasoning and behavior by altering the memory processes through which politically relevant considerations are retrieved.

When called on to express a judgment, given sufficient time to think and the motivation to query memory, the considerations that make their way into the conscious decision stream will be biased. As we will show in upcoming chapters, what comes to mind when voicing an opinion – as when responding to an NES open-ended question asking for likes and dislikes of the candidates and parties – will reflect what information is readily accessible from long-term memory, and this accessibility is strongly influenced by affective congruence.

Information in LTM that is congruent with the valence of current feelings will be favored in memory retrieval while affectively incongruent information will be inhibited. This means that the sampling of new considerations from LTM will generally support initial affective reactions, and as currently experienced feelings strengthen, the set of considerations retrieved into conscious WM will likely become more one-sided. In this way, reasoning processes that may seem to the citizen to provide reasons for one's evaluative reactions may more often rationalize the initial affect one felt toward the object of evaluation. This “snowballing” of affect over time will be most pronounced for people who have strong, univalent affect tagged to the object of thought and who have a substantial number of congruent considerations stored in LTM. In the political realm, it may be knowledgeable partisans who best fit this description.

As with affect transfer, affect contagion may be the product of extrinsic or intrinsic feelings. That is, affective tags already associated with political candidates, parties, issues, or other objects of thought – prior attitudes – will bias the retrieval of considerations and thereby alter the character of thought. When one is asked to produce and justify an opinion on affirmative action, for example, the prior attitude stored in affective tags linked to the affirmative action node in LTM will inevitably (and appropriately) color the response. One may have a variety of pro and con considerations on affirmative action, but the ones that are congruent with prior affect will more likely be retrieved. Such affect contagion is intrinsic to the object of thought. Unrelated or extrinsic feelings

can also bias the retrieval of considerations, using the same affect contagion processes. A bad mood, for example, or an unrelated negative prime (perhaps the word RATS subliminally presented in a campaign ad) will influence one's responses to the affirmative action question, reliably leading to the retrieval of more negatively valenced considerations.

We will show in upcoming empirical tests evidence of the systematic biasing effects of affective contagion on how citizens evaluate political leaders, groups, and issues, and how they judge policy proposals.

Given the seven postulates operating on a network of cognitive and affective associations, our expectation is that most citizens most of the time will be *motivated reasoners* who find it difficult if not impossible to evaluate attitude-relevant information in an evenhanded way. Challenged by attitudinally incongruent information, people will routinely rationalize the facts, figures, and arguments that they cannot effortlessly discount, depreciate, denigrate, or deny. Like the Bush 43 administration, citizens are prone in their everyday lives to fit the facts to their feelings. This is what we find in a series of experiments (reported in Chapters 4 and 5) exploring the impact of affect on political information processing. People find it difficult to override the spontaneously generated feelings triggered by their prior attitudes when evaluating political candidates, groups, and issues.

Forewarned Is Forearmed: General Expectations and Anticipated Objections

The simple act of evaluating is a human universal, “with survival depending on appropriately discriminating beneficial from harmful stimuli” (Ito and Cacioppo, 2005: 20). In our *John Q. Public* theory, as in Cacioppo, Gardner, and Berntson's (1997) *Evaluative Space Model* and Zajonc's (1980, 1984, 2000) *Primacy of Affect Model*, as well as the Marcus, Neuman, and MacKuen (2000) *Affective Intelligence Model*, evaluations are said to be generated by separate, somewhat independent positive and negative affective systems, each with unique activation functions that automatically adjust to fit the environmental context.

To the extent that *JQP* provides a reasonable account of the structures and processes of mental representations, most of our everyday life operates automatically (Bargh, 1997, 2007). The chapters that follow will document preconscious priming effects on political judgments, evaluations, and behavior, as well as even stronger effects of automatic affect when citizens are called upon to stop and think hard when evaluating political candidates, arguments, and policy recommendations. Virtually all the studies to be reported, and many more we will cite from the social and neuropsychological literatures, demonstrate the immediate or longer term consequences of unconsciously perceived affective events for sociopolitical evaluations and behavior. Before turning to the empirical tests of our theory, however, let us briefly list our major

hypotheses and then anticipate concerns about the construct validity, reliability, and explanatory power of implicit measures of attitudes.

General Hypotheses. At the core of this book stand three basic expectations: *feelings drive thinking more than vice versa; conscious experience always follows and is a product of unconscious processing; and behavior is often propelled by feelings through processes we do not consciously control.* It is our claim that conventional models of political behavior have the causal order wrong. Conventional models fail to appreciate the significance of information processes that occur on a millisecond timescale. And conventional models err in seeing citizens as (imperfectly) rational actors who construct preferences through conscious deliberation about the attributes of political actors, groups, or ideas.

Stated more formally, the major hypotheses we test in this book are:

- The *hot cognition hypothesis*, that all political objects that have been thought about in the past are tagged to positive and/or negative feelings.
- The *automaticity hypothesis*, that significant information processes occur outside of conscious awareness with substantial effect on subsequent conscious thought and behavior.
- The *affect transfer hypothesis*, that current affective states or primes, whether intrinsic or irrelevant, can transfer positive and/or negative feelings to objects of current thought. This can happen when affect and objects are consciously experienced or when one or both are outside of awareness.
- The *affect contagion hypothesis*, that affective states or primes, whether intrinsic or irrelevant, can influence the retrieval of considerations from memory, favoring thoughts with congruent over incongruent valence. This can happen for both consciously experienced affective states or when operating below conscious awareness.
- The *motivated reasoning hypotheses*, that prior affect will bias attention to and processing of information in ways that favor acceptance of affectively congruent arguments or evidence and rejection of incongruent information. This can occur for both consciously and unconsciously experienced affect.

In addition, we will test a variety of subsidiary hypotheses that will examine related processes, moderators, and mediators. Among these will be a chapter extending the hot cognition hypothesis to group identifications and racial attitudes, several studies examining the “snowball” hypothesis that consciously thinking harder *increases* the impact of unconscious affect, and research showing that motivated biases are greatest for sophisticates and those who care the most about politics. Notably absent from the empirical work we will present are two of our seven postulates: We will not present tests of the somatic marker hypothesis and the online processing mechanism for affect updating. The former has been extensively tested by others, though not in a political context (but see Morris, Squires, Taber, and Lodge, 2003), while the latter is well-established in the social and political psychology literatures, including some of

our own earlier work (Lodge, McGraw, and Stroh, 1989; Lodge, Steenbergen, and Brau, 1995).

We turn now to potential objections.

What are Implicit Attitudes and How Do They Relate to Explicit Attitudes?

Much ink has been spilt over the past fifty years in arguments as to how, when, and why within-individual attitude measures fail to cohere, with correlations routinely in the .30 to .50 range. These explicit attitude measures, moreover, routinely account for just 10 to 15 percent of variance in observed behavior, casting doubt on one of the pillars of modern social psychology, the notion that attitudes cause behavior. Not much to show for fifty-plus years of empirical research! A similar concern pertains to the *relationships among implicit and explicit measures of attitudes* and their predictions of behavior. Some readers may wonder what implicit attitudes really are and how they relate both empirically and theoretically to the more familiar self-reported attitude.

Many of the same faults, findings, arguments, and explanations surrounding the instability of explicit measures and their inability to better predict attitude-relevant behaviors also characterize contemporary research looking at the relationships among different implicit measures, between implicit and explicit measures, and implicit versus explicit predictions of behavior (Petty, Fazio, and Briñol, 2009). We have already noted that latent variable structural models analyzing implicit and explicit measures of attitudes toward social objects demonstrate that a two-factor model, with implicit and explicit attitudes as separate factors, is superior to a single-factor specification (Nosek and Smyth, 2007).

Meta analyses comparing implicit to explicit relationships over a large number of separate studies report (here we go again) correlations dispersed around .30 (Blair, 2001; Dovidio, Kawakami, and Beach, 2001; Greenwald and Banaji, 1995; Hofmann, Gawronski, Gschwendner, Le and Schmitt, 2005; Nosek, Greenwald, and Banaji, 2005). For example, Greenwald, Poehlman, Uhlmann, and Banaji (2009) compared implicit to explicit measures drawn from 184 independent samples and 14,900 experimental subjects, finding that correlations vary widely (from .18 to .68), with the average correlation for attitudinal, judgmental, and behavioral measures just .274. In the Nosek and Smyth (2007) review of 126 studies comparing Implicit Association Test measures to explicit like-dislike measures across a broad range of attitudes, correlations ranged from a low of -0.05 for reactions to Male-Female body images to a high of .70 for Pro Choice-Pro Life attitudes. For Republican and Democratic Party identifications, they found a healthy .59 correlation, while for Liberal-Conservative attitudes the correlation was .56. It would seem that one can find empirical cause to believe that implicit and explicit attitudes are the same, are related, or are utterly different constructs, depending on the type of attitude, context, or measurement strategy. Clearly, reporting the average coefficients of implicit-explicit consistency across many studies as some have done does not capture the complexity of these findings. Rather, as is also true for the relationships

among explicit attitude measures, the high variability in correlations suggests that moderator variables determine when correlations between implicit and explicit responses will be high, middling, or low.

As expected, given the motivation to control social impressions, the weakest correlations are uncovered when predicting explicit attitudes from implicit measures on such historically discriminatory attitudes as race, gender, and age, for such stigmatized people as the mentally and physically disabled, thin versus fat people, and those with AIDS, and for such risky behaviors as cigarette smoking, drug use, and unsafe sexual behavior. This finding suggests that implicit measures can sometimes reveal underlying attitudes that are not socially acceptable and therefore may not be revealed in overt surveys. In this sense, implicit measures may provide a “bona fide pipeline” to true attitudes (Fazio, Jackson, Dunton and Williams, 1995).

Before turning to a more detailed discussion of what are proving to be the most influential moderators of implicit-explicit relationships, let us note that *JQP*, as with virtually all modern-day models of evaluation in psychology, takes the “attitude-as-construction” perspective. When asked to think about why they prefer one object over another, respondents routinely construct their attitudinal responses on the fly, based on thoughts, feelings, and actions that are immediately accessible and easily verbalized (Wilson, Hodges, and LaFleur, 1995). People are found to construct their attitudes from:

- current thoughts and feelings (Chaiken and Yates, 1985; Judd and Lusk, 1984; Miller and Tesser, 1986; Wilson and Hodges, 1992);
- their present mood (Forgas, 1995; Petty, Schumann, Richman, and Strathman, 1993; Schwarz and Clore, 1983);
- their own behavior (Bem, 1967); and
- the immediate social context (Feldman and Lynch, 1988; McGuire, Padawer-Singer, 1976; Schuman and Presser, 1981).

The upshot of this view of attitudes-as-constructions is the expectation of attitude instability (which *JQP* shares with Zaller and Feldman’s 1992 survey response model). The weak to moderate correlations that characterize the relationships among explicit measures, among implicit measures, between explicit and implicit measures, and between attitude measures and behavior, is an everyday consequence of the attitude construction process. That said, well-known factors moderate relationships among and between attitudes and their measures. Research on moderators of implicit-explicit (I-E) relationships focuses on interpersonal and intrapersonal contexts and measurement effects.

As we have implied, the historical impetus for development of implicit measures was initially concern with the biasing effect of *social desirability* on direct self-report measures of personally and socially sensitive issues, with much of the research in the United States focused on “impression management” in racial, religious, and gender stereotyping. By minimizing the opportunity for strategic responding, implicit measures reduce the opportunity for respondents to guide

their responses. Nonetheless, self-presentation – that is, altering a response for personal or social purposes – does indeed moderate I-E correlations, almost always showing weaker correlations when social desirability is a concern for explicit measures.

Meta analyses find the *dimensionality of attitudes* to be an especially strong moderator of implicit and explicit measures of attitude consistency. Evaluations that conform to a simple, bipolar structure, in which liking for a concept (for example, pro-choice) implies disliking of a second concept (pro-life), tend to elicit stronger I-E correlations as well as increase the speed, consistency, and efficiency of processing – all hallmarks of automaticity.

A variety of internal factors tend to crystallize the evaluative dimension and increase I-E correlations. These include the number and quality of personal experiences in a particular domain (as is characteristic of political activists), identity-related comparisons (for example, male/female; American/foreigner; Republican/Democrat), and the frequency with which a response has been activated in the past. All of these factors promote the accessibility and consistency of attitudes toward a given object and consequently their automatic activation. A consistent finding across studies is that strong, well-defined attitudes elicit stronger I-E correlations than ones that are novel, unimportant, ambivalent, or infrequently thought about. Together, these intrapersonal factors account for a significant portion of variation in I-E correlations across domains, especially when the implicit and explicit attitudes are operationally measured by such simple affective ratings as good-bad, strong-weak, like-dislike, or warm-cold thermometer ratings (Hofmann, Gawronski, and Gschwendner, Le and Schmitt, 2005).

When individuals are *unable or unmotivated to search memory* or integrate additional information into an explicit evaluation, I-E correlations are typically stronger than when respondents are encouraged to deliberate before responding (Fazio, 1990). Three common ways of limiting cognitive processing is to encourage rapid responses, distract, or otherwise introduce additional cognitive demands so as to impede search and deliberation. There is a good deal of evidence showing that speeded responses and deliberative responses load on separate factors. Essentially, the richer the context, more complicated the format, or harder the questions, the more likely it is that *less-important* considerations will enter the evaluation. Here, we think the best course of action is to *not* follow Ben Franklin's dictum encouraging people to stop, think, and compute a preference, but rather to take the advice of Zen Master Chogyam Trungpa (1983): "First Thought. Best Thought."

Implicit-explicit consistency is also found to increase *when implicit and explicit attitudes are assessed similarly*. For example, following good measurement theory, modern explicit attitude measures often infer attitudes by aggregating levels of agreement across a variety of propositional statements, as does, for example, the Modern Racism Scale and the multiple item NES abortion scale. But the richness of these explicit measures may actually be a hindrance

to observing strong I-E consistency because implicit measures are found to reflect relatively simple good-bad associative relations as in the association of racial groups with positivity or negativity rather than nuanced measures of support or disapproval. The more dimensions, propositions, or challenges to an explicit attitude beyond the basic good-bad, like-dislike association, the more likely it is that implicit and explicit representations will diverge. If we are right in believing that online tallies represent the experienced costs and benefits of earlier evaluations, then relying on this implicit response would be both a reasonable and efficient strategy.

Implicit measures are sensitive to many and perhaps many more of the same *contextual factors* that impact explicit measures (Deutsch and Strack, 2010; Gawronski and Srithanan, 2010; Petty and Briñol, 2010). Initial evidence for the influence of context on implicit-explicit consistency is seen in a study by Wittenbrink, Judd, and Park (2001). Before completing several explicit measures of racial attitudes and an implicit Black/White subliminal priming measure of racial stereotyping, participants viewed a video clip of African Americans either in a gang-related urban setting or at an outdoor, suburban barbecue. Positive I-E correlations emerged only for participants in the gang-related video condition. For these respondents, their implicit stereotypic negativity toward African Americans was momentarily deactivated in the positive context.

While meta analysis of I-E relations shows that implicit and explicit measures are generally related, a positive correlation is of course no guarantee that the relationship isn't spurious. Construct validity is judged by how well the measures predict and explain a relevant behavior. Where, when, and the extent to which the measures diverge is difficult to gauge in the abstract. Of special note here, we once again call your attention to the interdependence of implicit and explicit processes and emphasize that automatic cognitive and affective processes *always* precede explicit responses. Given that implicit attitudes directly influence explicit attitudes, it is no wonder that the correlation between measures is typically positive, unless for strategic reasons respondents censure their immediate response or – this we think common – a citizen may have an implicit but not an explicit attitude toward the object.

In general, attitudes that are strong, important, certain, univalent, and have been evaluated frequently in the past yield higher I-E consistency than do attitudes that are weak, unimportant, uncertain, multipolar, ambivalent, or ephemeral. This pattern of findings suggests that I-E consistency is a function of how attitude objects are structurally represented in memory, in particular the strength of node-link associations, which in turn is a function of one's prior experience with the attitude object (Fazio and Zanna, 1978a, 1978b). A reasonable theoretical shorthand: implicit measures tap automatic and direct associations (i.e., OL tags) while explicit measures tap an integration of multiple direct and secondary associations from memory. The more complex the explicit measures, the more various will be the considerations brought to bear and the less the response will correlate with direct automatic associations.

Are Priming Effects Short Lived? Research by psychologists measuring the persistence of priming as well as framing effects on memory and attitudes is typically restricted to one-session studies, with estimates thereby limited to mere minutes. Perhaps the strongest tests of unconscious long-term effects are brought together in Merikle and Daneman's (1998) meta-analysis of priming on patients under general anesthesia. Across multiple studies they find evidence of the persistence of priming effects upwards of 36 hours, and in some studies cannot rule out consequences lasting three weeks to four months. In our single-session lab studies, we routinely find subliminal priming effects on $t_1 - t_2$ attitude change upwards of fifty minutes. Obviously, much more research needs to be done on the impact of *both* implicit and explicit information processing on everyday functioning. Consumer research seems the ideal domain for such studies, as experimental manipulations, for example, price, the color of packaging, or an athlete's endorsement can be randomly assigned at different locations, and the number of products sold provides a clear, easily measured effect. Though much more still needs to be done before we have a satisfactory answer to this question, such experiments in consumer research do show strong effects of unnoticed factors on purchases minutes to hours later (Maison, Greenwald, and Bruin, 2004; Mast and Zaltman, 2005).

External Validity. One might, and we know many colleagues do, object that experimental findings demonstrating the spontaneous impact of unnoticed priming events on beliefs and attitudes in the lab may be nothing more than a clever parlor room trick that only works (presumably with small effects) in a contrived experimental setting devoid of any "real world" complexity. True enough, our research, like most social-psychological studies of information processing, is based largely on research conducted in well-controlled, distraction-free settings, tapping behaviors free of immediate, serious consequences to life and limb. As is the case for all experimental *and* survey research there is a tradeoff pitting internal against external validity and it is always the case in the social and behavioral sciences that betwixt the two is an abyss few of us have seen across. In our studies, we opt to maximize internal validity, because if the internal validity of a survey or experiment is compromised, questioning the external validity would be moot. Our basic argument for taking our lab results seriously is that the capacity for unconscious processing is hard wired in the brain and our lab results are demonstrations of an effect that operates in real, artificial, and virtual worlds.

That said, it is certainly the case that not nearly enough work has been done to establish the validity of *either* laboratory or survey research on the expression of beliefs and attitudes in real-world settings, but there are numerous studies that do address the challenge (Perugini, Richetin, and Zogmaister, 2010). One of the first field experiments on political behavior, Gosnell's 1927 classic *Getting Out the Vote*, showed that simply asking citizens whether they expected to vote increased voter turnout, this a "mere measurement" effect that has been replicated many times over (Greenwald et al., 1987). One of the

earliest nonclinical studies we are aware of that looked at *unconscious* influences on attitudes was Razran's (1938) "luncheon technique" experiments, in which wall posters supporting different political causes were periodically displayed in a school cafeteria. Razran had people evaluate numerous political proposals before and after exposure to the posters, finding that although the participants were unable to discriminate those issues that had been exhibited on the lunchroom walls from those that had not, the proposals that had been displayed periodically at meal times gained significant approval. Razran, a student of Pavlov, interpreted this Type 2 cueing effect in terms of classical conditioning: food, a powerful unconditioned stimulus (the experiment was run at the height of the Great Depression), transferred affect to a paired object, here the unconsciously associated political issues.

Noticed but unappreciated effects are also at work in college classrooms where estimates of a teacher's height grow as a function of ascribed academic status (Wilson, 1968). And in the home, prerecorded laugh tracks on TV sitcoms, which most people say they dislike and claim have no effect, are shown to enhance the expressed enjoyment of shows (Fuller and Sheehy-Skeffington, 1974). More serious real-world social behaviors show even stronger effects, with a large literature showing a significant increase in suicides and suspicious single-driver vehicle fatalities following mass media reports of suicides, especially when the death is a celebrity of the same gender and age (Phillips, 1979; Gould and Schaffer, 1986). What is most striking about these real-world demonstrations of unconscious influences on behavior is that the priming effects appear stronger outside than inside the hallowed walls of the laboratory (Bushman and Anderson, 1998).

Much of the contemporary research on the external validity question is focused on the pursuit of goals, with many studies demonstrating that an intention to act can be as easily activated, as are semantic and affective associations (Bargh et al. 1996; Ferguson and Porter, 2010). The focus on goal behavior provides a compelling test of unconscious priming effects as the initiation and pursuit of goals is thought to be the epitome of volitional behavior (Chartrand and Bargh, 2002; Elliot and Fryer, 2007). Recall how the simple priming of the concept elderly led experimental participants to walk more slowly to the elevator and how the display of business paraphernalia promoted competitive behavior. In these studies not only were participants unaware that a goal had been primed but they also had no awareness of how or why they "consciously" decided to pursue the goal. Other research shows influences on goal-directed behavior generated by the mere passive activation of such relevant mental concepts as intelligence, politeness, power, cooperation, and achievement (Chen, Lee-Chai, and Bargh 2001).

Why Focus on Affect Rather Than Emotion? John.Q.Public is an affect-driven model and throughout we will have little to say about discrete or dimensional emotional responses, despite the fact that much recent research

in political psychology focuses on its diffuse form, mood (Forgas, 2000), or on such specific expressions of emotion as anger, anxiety, fear, and hope (Brader, 2005, 2011; Huddy, Feldman, Taber, and Lahav, 2005; Just, Crigler, and Belt, 2007; Marcus, Neuman, and MacKuen, 2000; Marcus, MacKuen, and Neuman, 2011; Small and Lerner, 2008; Valentino, Hutchings, Banks, and Davis, 2008). In contrast, our approach focuses on the initial, more basic effects of positive and/or negative affect on the expression of beliefs, attitudes, and behavior. As we theorized above, valence affect is primary in setting the direction of response, in guiding subsequent appraisals of the situation, and in determining when and how citizens may or may not label their feelings as an emotion.

William James (1884) proposed what is still today one of the more compelling ideas on the psychology of emotion; to wit: each of the specific emotions is characterized by a unique pattern of somato-vascular changes, and it is the perception of these bodily changes that differentiates one emotion from another. Despite hundreds if not thousands of psycho-physiological studies testing this specificity of emotions hypothesis, it may come as a surprise to learn that while there are many studies showing specific patterns of autonomic nervous system (ANS) activity for one or more of the emotions (Ekman, 2003; Ito, 2010; Lerner and Keltner, 2000; Lerner, Small, and Loewenstein, 2004; Panksepp, 1998; Scherer, Shorr, and Johnstone, 2001), there is equal if not more disconfirming evidence (Barrett, 2006; Levenson, 2003; Russell, Bachorowski, and Fernández-Dols, 2003; Turner and Ortony, 1992). Although we often find some mean differences in such ANS responses as blood pressure, muscle tension, and sweat gland activity for a discrete emotion, most studies report modest to weak correlations, with still lower correlations across seemingly related situations (see the meta-analysis by Barrett, 2006). Heterogeneity trumps uniqueness (Zajonc and McIntosh, 1992).

Levenson (2003), in his overview in the *Handbook of Affective Sciences* on the psychophysiology of ANS patterns of discrete emotions, summarizes the current state of affairs:

Even if emotion elicitation tasks were usually successful in producing the desired emotion in most participants; even if the autonomic nervous system was inactive before and after being recruited in the service of emotion; even if emotion solicitations in the laboratory had the kind of sharp onset, close match to prototype, and high intensity that reliably produced full-blown emotional reactions; even if the autonomic concomitants of specific emotions were dramatically different; and even if effect sizes were huge, then it would still be critical to ensure that the autonomic physiology derived on a particular trial from a particular participant was in fact associated with the actual occurrence of the targeted emotion. In reality, none of these “ideal case” scenarios is likely to be true. Even the best of the available elicitation tasks often have unintended emotional outcomes; the autonomic nervous system is continually acting in the service of many masters other than emotion: laboratory-induced emotional elicitations are often

pale comparisons of real-life ones; participants' emotional responses are often of low intensity and often include emotions other than the intended one; autonomic correlates of emotions are not unique but rather show complex patterns of overlap; and effect sizes are small.

What *can* be reliably differentiated by ANS measures is positive from negative affect (Lang et al., 1993; Ito and Cacioppo, 2005; Russell, 2003; Russell and Barrett, 1999).

Emotions, like beliefs and attitudes, but unlike valence affects, are constructed from what is accessible in memory, whether consciously perceived or not (Ruys and Stapel, 2008; Winkielman and Berridge, 2004), and in *JQP* the online tally representing the rewards and punishments of past evaluative experience anchors the construction process. A defining characteristic of emotional experience is its context, which triggers the reaction and gives feelings their shape, meaning, and functionality. The importance of the situational context for the labeling of an emotion is made apparent in studies that show how easy it is to manipulate how people categorize their emotions through subtle, even subliminal primes, or by varying contextual factors of which the person is unaware, hence the steep forgetting curve for source and contextual factors. You may remember having felt elated when your candidate scored a point in a debate or angry at the opposition's tactics, but you cannot reexperience the feeling itself (Niedenthal and Showers, 1991; Robinson and Clore, 2000). Except perhaps for strong, personally engaging experiences that were accompanied by heart-felt bodily responses (what Damasio, 2010, calls "body-loop" experiences), people have poor conscious access to such "objective" indicators of emotional experience as their heart rate, blood pressure, and other somatic changes, which could serve as important signals for labeling the emotional experience beyond good vs. bad and arousal (Cunningham and Van Bavel, 2009; Schachter and Singer, 1962; Strack, Martin, and Stepper, 1988). Note that somatic markers send simple signals of positivity-negativity and arousal that point out the direction of emotional response but cannot under everyday circumstances discriminate among the various negative or positive emotions.

The advent of brain imaging technologies has led many scientists to search for emotion specificity in the form of localization of discrete emotions. Two recent meta analyses of PET and MRI studies testing the hypothesis that fear, anger, sadness, disgust, and happiness have distinct neural circuits failed to find strong, consistent, or unambiguous evidence of localization in the brain for the specific emotions other than amygdala activation for fear (Murphy, Nimmo-Smith, and Lawrence, 2003; Phan, Wager, Taylor, and Liberzon, 2002). Moreover, thousands of connections to virtually every other brain module make it difficult to find a signature pattern for a specific emotion (Ito, 2010). Perhaps bigger and faster magnets with better spatial and temporal resolution will find

localized emotions (Fowler and Schreiber, 2008; Ito, 2010; Westen, Blagov, Harenski, Kilts, and Hamann, 2006), but we are not there yet (van Veen, Krug, Schooler, and Carter, 2009).

Note that fMRI imaging cannot test the hot cognition, primacy of affect, or affective contagion postulates central to *JQP* because the fMRI brain-image “slices” are recorded in seconds while affect operates on a timescale of milliseconds. Electroencephalographic (EEG) recordings of brain wave activity from the scalp can be used to measure the activation of affect and its effects on cognitive connections, as in the Morris, Squires, Taber, and Lodge (2003) test of the hot cognition hypothesis, but cannot clearly identify spatial signature patterns.

Problems are apparent on the behavioral side as well where emotional responses are found to vary significantly by whether or not the experimental/survey setting allows the individual to express an appropriate coping strategy. For example, being angry may lead you to yell, stomp your feet, lash out, sit back seething, or walk away. The expression of anger is dependent in part on whether the antagonist is a powerful bully, your boss, an underling, a politician unreachable behind the TV screen, or something as uncontrollable as 9/11. Fear can be expressed behaviorally by vigilance, fleeing, or freezing like the poor soul in Edvard Munch’s *Scream*. William James (1884) was right in believing that “thinking is for doing.” Emotions evolved as behavioral triggers and coping mechanisms, yet most social science studies do not allow the individual to act out, to strike out, cower, vent their anger, or express any other “hot” emotion. This is a serious problem for political science research on emotion, where most of our studies look at very tempered responses of tepid emotions in inconsequential settings.

Parental Warning: Consider this “hot” example of the motivating power of arousal (and of the weak correspondence between survey response and real-world behavior). Ariely and Loewenstein (2006) asked how the sexual attitudes and behavioral intentions of rational, intelligent people (as a proxy, Berkeley male heterosexual undergraduates) change from a “cool” survey setting to when they are in an impassioned state. Twenty-nine questions were asked twice, the first time in a survey setting, the second time in the participant’s dorm room, where alone with a copy of *Playboy* and a Saran-wrapped computer they recorded their “Yes”/“No” answers.

One series of questions asked the men to rate the attractiveness of different sexual prospects, among them (with first “cool” then “hot” Yes percentages in parentheses):

- “Can you imagine being attracted to a twelve-year-old girl?” (23 × 46%);
- “Can you imagine having sex with a sixty-year-old woman?” (7 × 23%);
- “If you were attracted to a woman and she proposed a threesome with a man would you do it?” (19 × 34%).

Other questions asked for the likelihood of engaging in such immoral acts as:

- “Would you keep trying to have sex after your date said no?” (20 × 45%);
- “Would you tell a woman that you loved her to increase the chance that she would have sex with you?” (30 × 51%);
- “Would you slip a woman a drug to increase the chance that she would have sex with you?” (5 × 26%).

In every case the young men’s answers were dramatically different in their aroused state. Unaroused, they do not know what they think, like, or will likely do when in an ardent state. Self-protection, mainstream sexual conservatism, even morality were swept aside. Ariely (2008: 97) sums it up thus:

When the participants were in a cold, rational, superego-driven state, they respected women; they were not particularly attracted to the odd sexual activities we asked them about; they always took the moral high ground; and they expected that they would always use a condom. They thought they understood themselves, their preferences, and what actions they were capable of. But as it turned out, they completely underestimated their reactions.

True enough, for many of us the everyday life of conventional politics is not as arousing. But should we not assume that the emotional states of anger, jealousy, and excitement are similarly affected? Compared to the true colors of emotional experience, the recollected and prospective response is a bland, “as-if” experience, what Niedenthal and her colleagues (Niedenthal, Halberstadt, and Innes-Ker, 1999; Niedenthal, Halberstadt, and Setterlund, 1997) describe as a “hot-cold empathy gap.” Predictions from emotion to behavior prove more reliable when the accompanying visceral response bolsters the attitudinal or behavioral response. The best chance for success in predicting public opinion and behavioral intentions are those experimental settings that manipulate emotions by exposing people to campaign ads, newscasts, or movie clips (Ansolabehere, Iyengar, Simon, and Valantino, 1994; Brader, 2006; Gilliam, Iyengar, Simon, and Wright, 1996; Valantino, Hutchings, and White, 2004), or challenge their beliefs and attitudes with real counterarguments and counterfactuals (Sniderman, Brody, and Tetlock, 1991; Taber and Lodge, 2006), or manipulate anxiety by having participants think about their own deaths and the rotting transformation of their bodies in the grave (Pyszczynski, Solomon, and Greenberg, 2003). Experiments that manipulate emotions grippingly appear better able to bridge the empathy gap by generating an experienced emotion rather than relying on an after-the-fact recollection.

Let us be clear: the question is *not* whether or not people have emotional experiences or whether they can read emotional expressions in themselves and in others. Surely they do, albeit not particularly well (Norris, Dumville, and Lacy, 2011). On this everyone is onboard. We all agree that great art, policy proposals, and politicians are more successful when they pull at the heart strings rather than appeal to the “brain” (Westen, 2008). What is problematic

is when the attitude is measured cold while the predicted behavior is hot. The key question is whether people can reliably discern among and between their emotions, say anger from fear, and whether researchers can reliably discriminate them. The literature is clear: the answer is: “yes,” “no,” “sometimes,” “to some degree or another” (Brader, 2011; Barrett, 2006; Ladd and Lenz, 2011; Marcus, MacKuen, and Neuman, 2011; Norris, Dumville, and Lacy, 2011; Rolls, 1999; Russell, 2003; Solomon, 2003). All that can be reliably discriminated from the correlate structure of neural, physiological, and subjective responses is the more basic good-bad affective response.

Moreover, it is typically the case that measured discrete emotions correlate so highly within the negative and positive domains (Barrett and Russell, 1998) that the specific emotions routinely fail to capture much unique variance (Feldman, 1995; Watson and Tellegen, 1985). The question emotion theorists need to answer is how much unexplained variance is left after controlling for valence? As we will show in the conclusion with NES data, not much.

Our view, shared with many appraisal theorists is that fundamental to all emotion responses is a common antecedent – positive and/or negative affect and arousal (see the overview by Forgas, 2003). Our focus on the primacy of affect fits comfortably within the affect-driven *primary appraisal stage* of appraisal theories of emotion in which an environmental situation is spontaneously perceived as positive, stressful, or irrelevant (Bower and Forgas, 2001; Clore, Schwarz, and Conway, 1994; Forgas, 2003; Niedenthal and Halberstadt, 2000). Generally, appraisal theory assumes that on exposure to an emotion-evoking stimulus, people *preconsciously* categorize their perception of the object as positive and/or negative. Critical here is that this immediate reaction occurs moments before a person is consciously aware of the stimulus, and may, *if* the individual is afforded the time and is sufficiently motivated, spur a conscious *secondary appraisal* seeking out reasons for “why I feel this way” (Robinson and Clore, 2002). The intensity of the affective response – presumably the degree of its sympathetic and parasympathetic activation – promotes a felt need to act in an affectively congruent manner, but the specific action taken depends on one’s unconscious and conscious appraisals of the triggering event in context, which in turn depends on what coping strategies exist.

As we see it, in general agreement with the classic Schachter and Singer (1962) theory of emotion, appraisal of an emotion’s cause produces the emotional label. Emotions are always about something; one cannot experience fear without being aware of the gun, be angry without seeing the opponent’s sneer, be depressed without seeing the pictures of starving children. Appraisal of the cause of an affect is what produces an emotion. The secondary, subjective report of an emotional experience is the postappraisal labeling of one’s affective response, which, if hot, is experienced viscerally, though perhaps not consciously (Lerner and Keltner, 2000). One may feel aroused, but the emotion itself requires a reason for the feeling. The actual label given the somatic (“gut”) experience is context dependent and culturally-based, and will

be strongly influenced by folkloric explanations for how people should respond in such situations (Robinson and Clore, 2000).

Again, we do not deny the existence of emotions or their demonstrable effects on animal and human behavior, but we are skeptical of our present-day skill as social scientists to reliably discriminate among like-valenced emotional responses (which routinely correlate in the .60–.80 range). There is incontrovertible evidence that a simple, good-bad, approach-avoidance response enters the decision stream spontaneously before an appraisal of the situation may guide the labeling of an emotion as anger, fear, anxiety, hope, or enthusiasm. Given the primacy of affect, once valence is controlled for there is precious little remaining unexplained variance for similarly valenced discrete emotional responses. The problem is exacerbated in the social sciences where our manipulations of emotions in experiments are typically weak. For example, we might induce sadness by having participants listen to a recording of the second movement of Schubert's *Death of a Maiden* while reading a news account of a candidate's policy proposals, or worse yet "manipulate" anxiety by informing respondents that a hypothetical candidate takes an issue position at odds with their own. At best such manipulations promote tepid, as-if emotional responses. The problem doesn't stop here: Few of our dependent variables capture an emotion as a coping mechanism by providing respondents the opportunity to strike out in anger, slump sadly, cringe, or actively seek out and cope with a threat.

Looking Ahead

The remainder of the book puts our key hypotheses to the test.

Chapter 3 will examine the *hot cognition hypothesis* with respect to political leaders, groups, and ideas. Chapter 4 will extend hot cognition to group and identity objects. In both chapters, we will seek to establish the *automaticity* of affective reactions to political stimuli. Chapter 5 will test the *affect transfer hypothesis* in the context of candidate evaluations, and will take a close look at our expectation that careful deliberation will increase the impact of unconscious processes. *Affect contagion* is the focus of Chapter 6, which will examine the degree to which the conscious generation of thoughts or considerations can be influenced by unnoticed affective primes, with downstream consequences for policy attitudes. Chapter 7 presents our empirical work on motivated reasoning about political policies. Our theory has been formalized as a computational model (*JQP*), and tests of the workings of this model are the subject of Chapter 8.

The seven postulates defining *JQP* place us squarely at odds with several of the most prominent models of public opinion, chief among them: (1) Zaller's (1992) "Receive-Accept-Sample" (RAS) model that informs his classic *The Nature and Origins of Mass Opinion* and the Zaller and Feldman (1992) application of the model to the survey response; (2) Marcus, Neuman, and

MacKuen's (2000) Affective Intelligence model; and (3) Lau and Redlawsk's (2006) models of heuristic decision making. These models are among the very best we political psychologists have developed to date and each makes a unique, positive contribution to our understanding of how citizens process information and inform their behavior. *JQP* shares much with each of these models at a general, descriptive level but differs significantly in theoretical process, in key predictions, in the manner of hypothesis testing, and in how we interpret the empirical results, both ours and theirs. We leave a discussion of the similarities and sharp differences between models to the conclusion in Chapter 9, when the reader will have the empirical tests of *JQP* in hand to make an informed evaluation of competing claims.