PREFERENCE FORMATION

James N. Druckman
Department of Political Science, University of Minnesota, Minneapolis, Minnesota 55455-0410; e-mail: druckman@polisci.umn.edu

Arthur Lupia
Department of Political Science, University of California, San Diego, La Jolla, California 92093-0521; e-mail: alupia@ucsd.edu

Key Words  belief formation, accessibility, on-line processing, preference reversals, persuasion

Abstract  This review concerns political preferences—what they are and where they come from. We begin by documenting the close relationship between processes of preference formation and change. Rather than suddenly appearing, most preferences emerge from interactions between individuals and their environment. This aspect of preference formation poses a concrete challenge: to uncover the mechanics of these interactions in important social contexts. We then describe political science research that meets this challenge. We find an expansive literature that clarifies how phenomena such as parties, campaigns, and the need to act strategically affect preferences. This work provides many widely applicable insights.

INTRODUCTION

Why do people do what they do? An easy answer is that people do what they want. But why do people want what they want? This question has fascinated intellectuals for centuries and continues to motivate researchers today. What scientists, generally, and political scientists, in particular, have learned about the answer is the topic of our review.

In what follows, we begin by defining what many scholars mean by “preference,” paying particular attention to what scholars outside of political science have discovered about the nature and origins of preferences. These discoveries clarify how individuals transform the “great, blooming, buzzing confusion” of human experience into preferences for objects such as candidates and policies (James 1890:488).

We then move to the core of the review: a description of political science research relevant to preference formation. This core has two sections. The first section focuses on the internal dynamics of preference formation. We compare and contrast some of political science’s most prominent approaches to explaining how
humans convert old and new information into political preferences. In the second section, we review work in political science that explores preference formation’s external dynamics. This research shows how occurrences outside the brain affect preferences inside it. For example, when some political actors have an incentive to provide information strategically, others may face changed incentives for interpreting the information they receive (i.e. they must choose whom to believe). When factors such as strategic behavior affect the quantity or character of available information, they can also cause people to form different preferences. The point of the research we review is to clarify the dynamics of such effects.

Although our review is expansive, it is far from comprehensive. The size of the literature relevant to preference formation necessitates a focused presentation. The literature is broad not only because of widespread interest in the topic of preference formation but also because questions of preference formation and preference change are, for most practical purposes, the same. This kinship is well established in cognitive science and psychology, where differences between the formation and change of related ideas (e.g. concepts, attitudes) are considered differences of degree rather than of kind. In the context of political science, this same kinship unites research on topics as diverse as on-line processing, persuasion, and strategic communication under the common domain of research relevant to preference formation. It also suggests new avenues for future research on preference formation.

THE NATURE AND ORIGINS OF HUMAN PREFERENCES

Theories of human cognition are ultimately theories of physical, biological systems. Our ability to describe human cognition in one way rather than another rests ultimately on the physical and biological nature of human beings. Furthermore, the fact that human beings are grounded in the world implies additional constraints that must be taken into account in constructing our theories. (Newell 1990:42)

To proceed to questions of preference formation, we need a working definition of “preference.” We say working definition because there is no global agreement on a single definition of the term. Indeed, terms such as preferences, tastes, and values are often used interchangeably throughout the social sciences. We define a preference as a comparative evaluation of (i.e. a ranking over) a set of objects. A preference serves as a cognitive marker that reminds people how to interact with various aspects of their environment. Preferences are stored in memory and drawn on when people make decisions. When, for example, people say that they prefer Lincoln to Douglass, they identify an aspect of their environment that, in their mind, provides them with greater benefits than other environmental aspects.

In this section, we begin our review of research on preference formation by looking at insights from cognitive science. We do so not because cognitive science has had a direct influence on political science—it has not—but because it helps
us to understand the nature and origins of preferences. We proceed by reviewing basic cognitive-science insights regarding the component parts of a preference—the objects of preference and comparative evaluations.

The Objects of Preference

The amoeba cannot choose whether to categorize, it just does. The same is true at every level of the animal world. Animals categorize food, predators, possible mates, members of their own species, and so on. How animals categorize depends upon their sensing apparatus and their ability to move themselves and to manipulate objects. (Lakoff & Johnson 1999:17)

The objects of preference are aspects of the environment that are evaluated relative to one another. They can include observable, physically continuous phenomena (such as bowling balls) and unobservable, physically discontinuous phenomena (such as shared ideas).

The objects within a preference are those that a person can imagine as substitutable. For example, shoes and tofu are substitutes within many domains, including apparel, where most people prefer shoes, and eating, where most people prefer tofu. If people cannot imagine what eating shoes or wearing tofu would be like, then they cannot have such preferences.

The predominant view of human cognition for nearly 2000 years has been that the objects of preference (alternatively, the categories of phenomena over which preferences can be held) are strictly external. In recent years, the evidence against this view of cognition has been piling high. Consider, for example, colors. Most people prefer some colors to others, and many have a favorite. People treat colors as basic attributes of other objects. But are colors just out there in the world waiting to be evaluated by the next sentient being who comes along? The answer is no. Colors, as we know them, are as much human creations as they are exogenous characteristics of our surroundings.

As Churchland & Sejnowski (1992:221) explain, “[c]olor perception depends on wavelength, but it is not identical to wavelength.” For example, a “red wagon may look red under a wide range of illuminations, including broad daylight, dusk, candle light, fluorescent light, and so forth, where the physical wavelengths actually impinging on the retina vary dramatically” (Churchland & Sejnowski 1992:221). It is more accurate to think of color perception as the product of an equilibrium between brain, body, and world. As Lakoff & Johnson (1999:23,25) explain,

[C]olors do not exist in the external world. Given the world, our bodies and brains have evolved to create color. Our experience of color is created by a combination of four factors: wavelengths of reflected light, lighting conditions, and two aspects of our bodies: (1) the three kinds of color cones in our retinas, which absorb light of long, medium, and short wavelengths, and (2) the complex neural circuitry connected to those cones....
meaning of the word *red* cannot be just the relation between the word and something in the world (say, a collection of wavelengths of light or a surface reflectance).... It must make reference to the neural circuitry.

That we can communicate preferences and other phenomena based on the color “red” is a testament to similarities in our cognitive and perceptive architecture more than to any universal property of surface reflectance. Indeed, two people whose retinas reacted very differently to light could find it difficult to communicate color-related concepts effectively.

An important implication of such research is that personal experience affects the set of objects over which people have preferences. For example, some people prefer dogs to cats as pets. Some make further distinctions between kinds of dogs and kinds of cats, whereas others are indifferent to dog or cat breeds. Indeed, the extent to which people hold preferences over types of objects depends on their prior experiences with related objects. For example, people who are very distant from politics may have only a general dislike for politicians, whereas those whose well-being requires interactions with different types of politicians hold preferences with finer distinctions.

In sum, the objects of preference are not simply out in the world waiting to be ranked. They are, instead, objects that our perceptive capacity allows us to differentiate, that our experience gives us an incentive to differentiate and recall, and that our cognitive capacity allows us to remember and evaluate.

**Where Comparative Evaluations Come From**

Nature ... is heavily bound by achieved solutions to previously encountered problems. As a result, new cognitive garments seldom are made of whole cloth; usually they comprise hastily tailored amendments to old structures and strategies. (Clark 1997:81)

We contend that evaluations are based, in part, on beliefs; that beliefs are the product of interactions between brain, body, and world; and that understanding these interactions is the key to answering questions about preference formation. Indeed, explaining preference formation is the process of explaining how beliefs and evaluations emerge from correlations between what people experience and what they feel.

Preferences are rankings derived from comparative evaluations that psychologists call attitudes. We follow O’Keefe (1990:18) in defining an attitude as “a person’s general evaluation of an object (where ‘object’ is understood in a broad sense, as encompassing persons, events, products, policies, institutions, and so on).” Attitudes are people’s orientations toward objects.

Most objects have more than one recognizable attribute. As a result, many scholars think of evaluations in spatial terms—as multidimensional orientations toward objects. Many people, for example, like Bill Clinton’s politics but not his
penchant for extramarital affairs. We say that such people are “close” to Clinton
on the political dimension, but “far” from him on the marital-ethics dimension.
Evaluations of multidimensional objects are based, in part, on evaluations of
object attributes (see e.g. Fishbein 1963). Each attribute evaluation, in turn, is
based on beliefs. A belief is a neurally held probability distribution over possible
characteristics of an object attribute (Churchland & Sejnowski 1992, Kandel et al
1995). For example, an attribute of certain types of skies (e.g. cloudy and dark) is
that they produce rain. People base preferences over objects (such as the beauty of
different types of skies or the best time for a picnic) on their beliefs about object
attributes, such as the likelihood of rain.
Beliefs about attributes, in turn, depend on information. Following convention
in cognitive science, we define information as any data potentially relevant to
interactions between the brain, the body and, in some cases, the environment.
Beliefs, therefore, are the result of interactions between brain, body, and world.
Learning is the process by which new information leads to a change in beliefs
about attributes of the object being studied. Churchland & Sejnowski (1992:117)
call learning “[A]dapting to the present on the evolutionary-tried assumption that
the future resembles the past.”
Having traced the ancestry of preference, we confront the question, “Where
do particular preferences come from?” The recent literature in cognitive science
suggests the following explanation.
From very early stages in fetal development, the brain receives information
from and sends instructions to other parts of the body. Among the most important
signals sent to the brain are those regarding pain. Put differently, “Biological
cognition is highly selective, and it can sensitize an organism to whatever (often
simple) parameters reliably specify states of affairs that matter to the specific life
form... [organisms] rely on simple cues that are specific to their needs, and... profit
by not bothering to represent other types in detail” (Clark 1997:25). Aspects of
eyearly brain-body communciqûes are stored in the brain and become the basis for
future instructions from the brain. Physically, much of this information is stored
as growth in dendrites. The resultant neural network formed by dendritic growth
then facilitates or inhibits the subsequent chemical and electrical transmissions
that serve as the medium for future brain-body communications.
As the fetus develops, early brain-body communications come to correlate with
variations in the fetal, and subsequently the infant, environment. Preferences for
or against certain environmental aspects arise. The fetus and infant come to prefer
(i.e. want to be near) the identifiable aspects of their environment that correlate
with an increase in pleasure or a decrease in pain. “[F]or an infant, the subjective
experience of affection is typically correlated with the sensory experience of
warmth, the warmth of being held.... Universal early experiences lead to universal
conflations...” (Lakoff & Johnson 1999:46).
As the infant learns that it can manipulate aspects of its environment (e.g. that
criying induces parental attentiveness and that food can be thrown), it realizes the
concept of choice. When sensorimotor skills develop, the infant has an incentive to compare a larger set of objects.

Early on, a child's preferences over most classes of objects are based on direct experiences. As Holland et al (1986:9) explain, the process “is (a) directed by problem solving activity and (b) based on feedback regarding the success or failure of predictions generated by the system.” As children gain the ability to communicate, however, they begin to experience objects indirectly. Through stories, for example, children develop affinities for abstractions such as Peter Pan and fears of abstractions such as monsters. When sensorimotor and communicative organs develop fully, the architecture for all subsequent preference formation is in place. People can then begin to develop preferences for objects as abstract as political parties and constitutional forms.

Preference Change and Preference Formation

Many of the preferences that adults hold are different from those they held as children. Interactions with their environment cause these changes. However, the number of interactions is far greater than the number of changes. When do preferences change? And how different are the mechanics of preference change from the mechanics of preference formation?

To answer these questions, it is important to note that “preference change” can mean either of two things. First, we can change preferences between two types of objects, $A$ and $B$, because we gain new information about the attributes of at least one type of object. (For example, when a child learns that certain dogs bite, her preference for playing with those dogs, relative to other things, may shift.) Second, we can decide that the objects formerly known as $P$ are better treated as objects of distinct types $P_1$ and $P_2$.

In both cases, new information about an object attribute causes a change in the comparative evaluations of a class of objects. Preferences over “new” objects are examples of the second kind of change. However, when we encounter stimuli that are novel to us, we instantly form beliefs about them. These beliefs—and the preferences that follow—are not newly created; rather, they emerge from preferences and beliefs we have about objects with which we have interacted in the past. The objects we select from the past are those that we perceive to be similar to the novel stimulus.

Holland et al (1986) clarify the conditions under which such changes are most likely and the consequences of such change. The literal process they describe consists of rules of induction, but these rules are equivalent to conditions for belief change.

“Two important kinds of triggering conditions are the failure of a prediction and the occurrence of some unusual event” (Holland et al 1986:80).

“A new rule is built from the existing one by (a) augmenting the condition of the existing rule with additional unusual properties of the failure context and
(b) substituting the unexpected outcome as the action of the new rule” (Holland et al 1986:88).

“Thus, competition will favor those rules that (a) provide a description of the current situation, (b) have a history of past usefulness to the system, (c) produce the greatest degree of completeness of description and (d) have the greatest compatibility with other currently active information” (Holland et al 1986:49).

Seen from this perspective, the mechanics of preference formation and preference change are indistinguishable for most adults. Explaining preference formation means explaining how beliefs and attitudes emerge from correlations between what we observe and what we feel. From early in life, however, novel stimuli are understood by their relationship to stimuli that are more familiar. As a result, there is no difference between preference formation and preference change. New preferences emerge from existing ones (e.g. we learn that objects of type \(P\) are better treated as objects of type \(P_1\) and \(P_2\)). “New” preferences are not created out of thin air; they evolve from refinements of existing preferences. Like many prominent psychological treatments of attitudes (e.g. Eagly & Chaiken 1993:219), this review treats the formation and the change of beliefs, evaluations, and preferences as identical.

Summary

The objects over which preferences are held are not necessarily external. Indeed, many of what we perceive to be objects are conceptual creations that result from an equilibrium between brain, body, and world. Preferences over classes of objects are rankings that are derived from evaluations, where evaluations depend on beliefs, and beliefs are the result of interactions between individuals and their surroundings.

This view of the cognitive underpinnings of preference formation is not without competitors or controversy (see e.g. Lakoff 1987; Clark 1993, 1997; McCauley 1996; Pinker 1997; Lakoff & Johnson 1999). It is, however, an increasingly dominant view that is well supported by empirical research on brain function and conceptual development in cognitive science, the neurosciences, and psychology.

Some observers object that this point of view poses the threat of subjectivism. If objects are not simply external and if preferences derive from previous associations, then perhaps everything about our world, including morality, is a matter of opinion. The evidence supporting such a conclusion, however, is slight. Researchers have discovered basic and universal properties of brain function and many higher-level behaviors (e.g. Kandel et al 1995). These similarities, combined with omnipresent forces of nature (e.g. laws of gravity), put an upper limit on the variance with which humans perceive the world and the differences in the preferences that they can hold. Such discoveries reinforce the foundation of systematic, generalizable, and scientific studies of preferences.
PREFERENCE FORMATION IN POLITICAL SCIENCE

Internal Processes

In this section, we describe attempts to characterize the internal process by which individuals convert information from their environment into evaluations of political objects. Because these attempts differ in their portrayals of how people treat new information, they generate different implications for how preferences form. This review is not comprehensive. Rather it focuses on ideas developed recently and cited frequently.

We begin by comparing and contrasting memory-based and on-line models of information processing. We end by discussing the implications of preference-reversal experiments for the more general study of preferences.

Memory-Based Models

Many social science models share the implicit assumption that preferences form through a memory-based process (Lodge et al 1990). The idea behind a memory-based model is that people base their evaluations on information that they retrieve from memory. For example, when called on to evaluate an object, people canvass their memories for information on object attributes and use what they find to form preferences.

In comprehensive memory-based models, an individual recalls all relevant information and integrates it into an overall evaluation. For example, when an individual receives information about a candidate (e.g. the candidate’s issue positions), she files it away in long-term memory. Later, she retrieves the candidate’s issue positions on a host of issues from long-term memory, weights the information according to its importance or relevance, and finally integrates it into an overall evaluation (see e.g. Fishbein & Ajzen 1975, Enelow & Hinich 1984).

Such models assume that people engage in an enormous amount of computation. This assumption contrasts with the widely held view of individuals—voters in particular—as having neither the motivation nor the ability to engage in an exhaustive memory search and weighting of evidence. Numerous scholars have attempted to rectify this apparent contradiction by offering memory-based models that assume less computation (e.g. van der Pligt & Eiser 1984).

For example, Kelley & Mirer (1974:574; also see Kelley 1983) argue that a voter forms his candidate preference by canvassing “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 net number of favorable attitudes, if there is such a candidate....” Using this model, Kelley & Mirer accurately predicted vote choice for approximately 88% of voters surveyed during presidential elections from 1952 through 1964.

More recent work presents actors who are even less computationally complex. Accessibility models, for example, portray people as basing preferences on
PREFERENCE FORMATION

small sets of considerations. Accessibility refers to the likelihood that a given consideration (or construct) will be retrieved from memory when forming a preference (e.g. Bruner 1957:135). A large literature shows that people base many judgments on the considerations that are most accessible in memory (e.g. Higgins & King 1981, Wyer & Srull 1989, Fazio 1995). For example, if economic considerations happen to be accessible in a voter’s mind, she will base her candidate preference on economic considerations; if, instead, foreign affairs considerations are accessible, then the voter will base her preference on foreign affairs considerations (Sherman et al 1990).

A prominent example of such a model in political science is that of Zaller (1992) and Zaller & Feldman (1992). Zaller argues that people form evaluations by “averaging across the considerations that are immediately salient or accessible to them” (1992:49), where a consideration is defined as “any reason that might induce an individual to decide a political issue one way or the other” (1992:40). In Zaller’s model, people base their preferences on whichever considerations happen to come to the “top of the head” (i.e. whichever considerations happen to be accessible).

Which considerations are likely to float to the top? Zaller (1992:48) contends that “the more recently a consideration has been called to mind or thought about, the less time it takes to retrieve that consideration or related considerations from memory and bring them to the top of the head for use.”

An important implication is that citizens are unlikely to have true attitudes (see e.g. Wilson & Hodges 1992, Zaller 1992:93, Kinder 1993:48; for an alternative view, see Kuklinski & Hurley 1996:134–38). In Zaller’s model, citizens’ preferences depend on whatever happens to be salient at the time of expression. For example, if an individual recently overheard a discussion about economic issues, then economic considerations may come to the top of the individual’s head, and as a result, her expressed preference will be based largely on considerations of economic issues.


1It is unclear when a consideration is sufficiently accessible for activation (see e.g. Fazio 1995:273). Accessibility is similar but not identical to Tversky & Kahneman’s (1973) availability heuristic (Higgins & King 1981:71).
On-Line Models
Like the accessibility model, the on-line model of evaluation recognizes that memory limitations prevent people from engaging in an exhaustive information search every time they form a preference. However, the on-line model portrays people as overcoming this limitation in a very different way (see e.g. Hastie & Park 1986, Bassili 1989).

Instead of basing an evaluation on whichever considerations happen to come to mind, the on-line model suggests that people form and maintain a running “evaluation counter” of certain objects. When an individual encounters new information about such objects, she brings an affect-laden “evaluation counter” (i.e. running tally) into working memory, updates it using the new information, and restores the counter to long-term memory.

An important aspect of this model is that, after updating her evaluation, the individual may forget the information that affected the evaluation. When asked to express their evaluation, people simply retrieve the evaluation counter without searching for the information on which it was based. Lodge et al (1989:401) explain that the result may be “that people can often tell you how much they like or dislike a book, movie, candidate, or policy [because they maintain a running evaluation] but not be able to recount the specific whys and wherefores for their overall evaluation...” This is in sharp contrast to memory-based models, in which individuals do not maintain a running evaluation counter and instead base their evaluations on whatever information they happen to remember.

Lodge and his colleagues have been at the forefront of developing an on-line model of candidate evaluation. Their experiments show that subjects who engage in on-line processing base their candidate evaluations on information that enters their evaluation counter (over time) rather than the bits of information that happen to be available in memory at the time the evaluation is rendered. For example, a pro-choice, tough-on-crime voter may receive campaign information that a candidate supports abortion rights and strict federal crime laws. As a result, the voter accesses and updates her on-line evaluation of the candidate in a favorable direction, and then quickly forgets her reasons for doing so while restoring the on-line evaluation in long-term memory. At a later time, when the voter needs to evaluate the candidate (e.g. cast a vote), she simply retrieves the positive on-line evaluation and thus offers a favorable candidate evaluation, even if she does not remember the candidate’s stances. Thus, there may be no relationship between what the voter remembers and whom the voter prefers, or the relationship may reflect post hoc rationalizations.

If people form their evaluations on-line, then researchers should not expect people to remember and report the reasons for their preferences. This has a number of implications (Lodge et al 1995, Lodge & McGraw 1995). First, if the on-line model is accurate, then we should be wary of using recall measures such as open-ended like-dislike questions and issue-position questions to understand the basis for citizens’ preferences. These recall information measures may have no relationship to actual preferences, or they may be rationalizations for preferences.
Indeed, Rahn et al. (1994b) argue that the likes and dislikes that Kelley & Mirer (1974) use to explain candidate preference are actually post hoc rationalizations for on-line evaluations formed over time (also see Lau 1982; Lodge & Stroh 1993:257–61; Lodge et al 1995:311, 319–20).

Second, if voters form evaluations on-line, then we cannot judge the impact of a campaign by assessing how much campaign material voters remember, since a voter may use campaign information to update her evaluation and then dispose of the information (Lodge et al 1995). Both these points contradict memory-based models, which contend that what an individual happens to remember forms the ingredients for what she prefers (Hastie & Park 1986).

The on-line model also suggests that individual preferences are less susceptible to sudden shifts (Krosnick & Brannon 1993:965) than accessibility models would imply. In sharp contrast to accessibility models, the on-line model largely maintains the traditional view of preferences as more stable phenomena (Zaller 1992:50).

Memory-Based Versus On-Line

When do people form evaluations on-line and when do they base their evaluations on what is available in memory? The most basic answer is that people form evaluations on-line when they believe that a judgment will probably be required later (Hastie & Park 1986:262). In contrast, when an individual encounters information without a specific processing goal, or if the goal is to remember as much as possible, then the individual will probably not form an on-line evaluation. Moreover, when an individual lacks the motivation for accuracy or the opportunity to engage in an extended memory search, she may simply base her evaluation on whatever information happens to be accessible (Fazio 1990).

Which type of processing takes place in political settings? Although both Zaller (1992:279) and Lodge (see McGraw et al 1990) acknowledge that people may use different types of processing at different times, they disagree on the prominence of each type of processing.2 Zaller (1992:279) argues that “the on-line model is inappropriate in the domain of political attitudes...” (also see Zaller 1992:50), whereas Lodge et al (1995:321) argue for the on-line model when they state, “we believe there are many circumstances (political campaigns being a case in point) under which memory does not play a critical mediating role” (also see Lodge 1995:119).

A possible explanation for the differences between Lodge’s and Zaller’s claims is that Lodge focuses on candidate evaluation whereas Zaller focuses on survey response more generally. In the case of candidate evaluation, people may anticipate that a judgment will be required later (i.e. they know that they will have to vote), and so they form on-line evaluations. In contrast, most people do not anticipate

---

2Zaller and Lodge also recognize that the on-line versus memory-based distinction is not always clear (see e.g. Hastie and Pennington 1989).
answering survey questions, and thus, they cannot access on-line evaluations when
a surveyor surprises them with a question (see Kinder 1998b:813–14).

There is also some evidence that on-line processing occurs more often among
political sophisticates, whereas memory-based processing occurs more often
among those with less political sophistication (McGraw et al 1990, McGraw &
show, however, that the effect of political sophistication is context-dependent.
They find that sophisticates typically engage in on-line candidate evaluation, re-
gardless of the context, whereas less sophisticated people engage in on-line candi-
date evaluation in less demanding contexts (e.g. candidates presenting information
sequentially) and memory-based evaluation in more demanding contexts (e.g. a
debate between two candidates).

Preference Reversals and Their Implications
We close this section with a brief description of, and comment on, the implications
of experimental results documenting preference reversals. A common assumption
in the social sciences, particularly in microeconomic models, is that the compar-
itive evaluations in a preference have at least two properties (Kreps 1990:19–22;
see e.g. Arrow 1951:12–14, Downs 1957:6, Riker & Ordeshook 1973:16, Morrow
1994:18). The first property is that an individual cannot strictly prefer Policy A to
Policy B while also strictly preferring Policy B to Policy A. The second property
is that preferences are transitive, so that if an individual prefers Policy A to Policy
B and Policy B to Policy C, then she must also prefer Policy A to Policy C. A third
commonly assumed property of preferences, which is related to the first property,
is that preferences are invariant, meaning that “different representations of the
same choice problem should yield the same preference” (Tversky & Kahneman

Whether and when preferences have these properties is a point of great con-
tention. Fueling this debate is the existence of experimental subjects whose
preferences violate transitivity or invariance assumptions (e.g. Tversky 1969,
Lichtenstein & Slovic 1971, Grether & Plott 1979, Kahneman & Tversky 1979,
Tversky & Kahneman 1987, Quattrone & Tversky 1988, Tversky & Thaler 1990,
Rabin 1998). Perhaps the most widely cited example is that of Tversky &
Kahneman (1981, 1987). These researchers show that when choosing between
risky prospects, people prefer risk-averse alternatives when the consequences of
their choices are framed in terms of gains, and they prefer risk-seeking alterna-
tives when equivalent consequences are framed in terms of losses. This work is
consistent with a large literature showing that people’s preferences, as expressed
in surveys, change with alterations in question wording, format, or placement
(e.g. Mueller 1973, Sullivan et al 1978, Schuman & Presser 1981, Brady &
Zaller 1992). These types of examples are often taken as evidence that preferences
are neither stable nor invariant (e.g. Bartels 1998).
Such “preference reversal” results have had an important impact on how social scientists think about choice. However, the implications of these results, particularly the results of Tversky & Kahneman, for the formal study of preferences are often overinterpreted. Tversky & Kahneman’s results provide clear examples of the fact that people’s preferences depend on the information that they are given. Describe abstract objects (e.g. policies regarding a distant, fictional population) in terms of deaths caused and people rank them in one manner. Describe equivalent objects in terms of lives saved and people may rank them differently. But such demonstrations do not imply that people cannot, or do not regularly, hold transitive and invariant preferences in many political contexts (Sniderman 2000).

The preference reversal results reveal a problem in a more basic—and often unstated—assumption in many rational choice models, particularly older ones. The assumption is that preferences are fixed and exogenous (e.g. changes in information do not alter preferences). Tversky & Kahneman’s results clearly show that such an assumption is not viable in many cases of substantive interest. This valuable insight complements Simon’s (1979, 1985) effort to make formal modeling more cognitive. However, their results do not show that people are incapable of holding transitive and invariant preferences either generally or in many of the substantive contexts that political scientists care about.

Indeed, there are all kinds of preferences that appear to be stable and invariant for most people, including favorite political party, favorite religion, and favorite sexual orientation. For the many substantive domains in which information changes do not induce preference change, no external validity is sacrificed by using the traditional modeling assumptions.

In sum, experimental work on the relationship between preferences and information has produced some of the social sciences’ most important recent findings. At a minimum, these findings have led scholars to ask more pointed and constructive questions about the role of preferences in social science models (see e.g. Lacy 1997).

**External Influences**

Attempts to change the beliefs and preferences of others are a primary means of political action. In this section, we review scholarship on how such attempts affect political preferences. We focus on two categories of such studies: studies of persuasion and studies of strategic communication.

The first category contains the literature on political persuasion. Research on persuasion is relevant to preference formation because the point of persuasion is to change another’s beliefs. Recent research clarifies the conditions under which attempts to change the beliefs of others succeed. The seconds category contains a literature whose logical foundations are game theoretic. Contributors to this literature clarify the mechanics of belief change in situations where people have an incentive to be strategic in what they say. Because it answers questions such as
“How do people choose whom to believe?”, this literature clarifies the extent to which preferences are affected by the actions of others.3

**Persuasion Studies**

Persuasion, defined as “human communication designed to influence others by modifying their beliefs, values, or attitudes” (Simons 1976:21), plays an important role in politics. As Mutz et al (1996a:1) explain, “Politics, at its core, is about persuasion. It hinges not just on whether citizens at any one moment in time tend to favor one side of an issue over another, but on the numbers of them that can be brought, when push comes to shove, from one side to the other or, indeed, induced to leave the sidelines in order to take a side.”


**Recipient Effects** Sniderman et al (1991:8) explain, “People make up their minds in different ways.” Theirs is among recent work that examines how individual differences affect the impact of new information on beliefs and evaluations. For example, some recipients are persuaded by messages that others find unpersuasive, and researchers of recipient effects want to know why.

Perhaps the most studied recipient characteristic is the amount of political information the recipient possesses, also known as the recipient’s political awareness, expertise, involvement, knowledge, or sophistication (see e.g. Luskin 1987, Krosnick 1990, Sniderman et al 1990, Zaller 1992, Krosnick & Brannon 1993, Delli Carpini & Keeter 1996, Popkin & Dimock 2000). Many political scientists working in this area agree that (a) for one person to change another’s beliefs, the information recipient must receive and accept the message (McGuire 1968, Zaller 1992); (b) the most aware individuals are more likely to receive new political information; and (c) the most aware individuals’ beliefs are less likely to be changed by new information. The basis for this last point is the finding that well-informed (or highly aware) people are more likely to carefully and critically evaluate the messages they receive (see e.g. Zaller 1992, McGraw & Hubbard 1996, Cobb & Kuklinski 1997).

---

3Because of space limitations, this chapter does not describe research on persuasion in particular substantive domains such as mass media, or research on socialization (e.g. by neighbors and parents). See Kinder (1998a,b) for such reviews.
An increasingly common finding in this type of work is that people who are moderate in their awareness/information scores are the most likely to change their beliefs. These peoples’ beliefs are thought to be more prone to change because they are more likely to receive new information than less aware persons and are more likely to be persuaded by new information than more aware persons. Zaller explains (1992:124) that “the relationship between awareness and attitude change may be nonmonotonic, that is, that persons at middle levels of awareness may be most likely to change” (also see Kinder 1998a:183–84, 1998b:812–15; Nelson et al 1997b:227).

The implications of this work for questions of preference formation are as follows. The best-informed and worst-informed persons in the population have stable beliefs and evaluations about a wide range of political objects. As a result, they have stable preferences as well. The preference stability of the best-informed persons comes from the fact that they are already very knowledgeable, which gives them the ability to generate internal counterarguments and limits the extent to which new information surprises them. The preference stability of the worst-informed persons comes from the fact that they receive little feedback about their current beliefs and, therefore, little stimulation for preference change. Other citizens’ preferences are more susceptible to change.

**Message Effects**

How the content of a message affects the preferences of a listener can be difficult to study. Consider, for example, an attempt to determine how certain presidential phrasings affect the extent to which people prefer a president’s actions. Such a determination requires researchers to somehow identify and separate all of the other environmental factors that could affect beliefs or preference changes. This task is not easy. As a result, such studies are not numerous, particularly in political science. However, some political scientists have designed clever ways to uncover message effects.

Cobb & Kuklinski (1997), for example, exposed experimental subjects to arguments about the North American Free Trade Agreement (NAFTA) and health care. For each issue, they varied the structure of the argument in multiple ways. In one of the variations, they exposed subjects to either an argument that was in favor of the policy (pro) or an argument that was against the policy (con). Cobb & Kuklinski (1997:115) find that “con arguments win the day.” This is consistent with their expectation (1997:91) that when “making political decisions, [people] place more emphasis on avoiding potential losses than on obtaining potential gains” (also see Lau 1985). The status quo bias they find resembles findings elsewhere in the social sciences (e.g. Samuelson & Zeckhauser 1988).

These findings have important implications for the politics of preference formation. Chief among these implications is the attractiveness of “going negative” as a political strategy (e.g. Ansolabehere & Iyengar 1995). If people are more concerned about losses than gains, or if negative information about a particular candidate helps citizens draw finer distinctions between competing candidates, then citizens have a greater incentive to attend to negative information, and information
providers have a greater incentive to supply negative information. Such incentives will change the stream of information available to many citizens in ways that increase the likelihood that negative information will be the platform from which preferences emerge.

Source Effects  Many political messages come from known sources, such as the President, a party leader, a prominent politician, a media figure, or an interest group. Unlike research on message effects, explorations of the dynamics of source effects are abundant. This work has demonstrated the effect of source characteristics such as the speaker’s trustworthiness (Popkin 1994:47), public approval (Page et al 1987, Mondak 1993:195), insider status (Carmines & Kuklinski 1990:248), accuracy and objectivity (Iyengar & Kinder 1985), expert status (Page et al 1987), likeability (O’Keefe 1990:107; also see Brady & Sniderman 1985), party reputation (Iyengar & Valentino 2000), and ideology (Zaller 1992:47).

Kuklinski & Hurley (1994, 1996) provide a simple and effective example of such research. The researchers gave experimental subjects the following passage:

We would like to get your reaction to a statement that [SPEAKER] recently made. He was quoted in the New York Times as saying that African-Americans must stop making excuses and rely much more on themselves to get ahead in society. Please indicate how much you agree or disagree with [SPEAKER]’s statement.

Respondents randomly received the statement either without attribution to a speaker or attributed to one of the following speakers: Ted Kennedy, George Bush, Clarence Thomas, or Jesse Jackson. Kuklinski & Hurley find that African-American respondents demonstrated significantly greater agreement with the statement when it was attributed to Jackson or Thomas than when it was attributed to Kennedy or Bush (or when no attribution was given). African-American respondents were much more persuaded when the statement came from an African-American source than when the same statement came from a Caucasian source. The ideology of the speaker had a far smaller effect.

Lupia & McCubbins (1998) use a series of models, laboratory experiments, and survey experiments to probe more deeply into how people acquire and process information in political settings. They prove that persuasion requires that the receiver perceive the speaker to be knowledgeable and to share her interests. The speaker’s actual knowledge and interests, factors employed in many studies of persuasion, are of secondary importance. This result implies that the receiver’s perception of the speaker’s knowledge and interests are the fundamental source effects. All other speaker attributes—such as a speaker’s attractiveness, party, race, likeability, ideology, or reputation—affect persuasion only if they affect the listener’s perception of the speaker’s knowledge or interests.

Lupia & McCubbins’ many experiments put their most critical findings to the test. In a survey experiment, for example, they ask respondents whether they think spending money to build prisons is a good or bad idea. Before posing the question, however, they expose many of the respondents to a positive or negative
endorsement by Rush Limbaugh or Phil Donahue. They find that respondents follow their endorser’s advice when they perceive the endorser both to be knowledgeable and to share their interests. Further, it is these two variables that mediate persuasion and not likability, partisanship, or ideology. This and related laboratory experiments provide strong support for the conclusion that a receiver’s perceptions of a speaker’s interests and knowledge are primary factors in determining whether phenomena such as speeches, advertisements, or threats from abroad will affect the preferences of their intended audiences.

**Strategic Communication Studies**

Outside of political science, most empirical studies of preferences take place in a laboratory environment. Many of these studies follow the rules of the experimental paradigm so dominant in psychology, in which the experimenter picks the stimulus and the sole point of the experiment is to measure the subject’s response. Although such studies are very informative, many political contexts differ in an important way.

In politics, many people provide information to others in an explicit attempt to change beliefs, attitudes, or preferences. When a speaker wants to express an idea, there are often multiple ways to frame the issue at hand and choices to be made about the amount of information to provide. If such speakers base their choice of what to say on an audience’s likely reaction, we say that the speakers have chosen their words strategically.

When speakers choose their words strategically, two things happen. First, the content and amount of available information can change. Second, recipients of such information may have an incentive to process it differently. As a result, when speakers choose their words strategically, it can affect the information received, and consequently the preferences held, by the recipients.

Untangling the implications of strategic behavior for preference formation is tricky. An emerging branch of scholarship in political science, however, has provided some insights. Much of this work follows from the advent of strategic communication models (see e.g. Spence 1974, Crawford & Sobel 1982). Scholars use these models to clarify how a speaker’s characteristics affect the persuasiveness of what she says. The value of the models lies in their ability to generate precise and general insights about preference formation.

In the last two decades, political scientists have introduced strategic communication models that provide precise insights about persuasion and preference formation in important political contexts (see Banks 1991 for a review).4 Calvert (1985), for example, offers one of the earliest signs of the influence of strategic

---

4 Most formal modelers rely on the maxim that preferences are fixed and exogenous. However, the preferences treated as such are used in a model to define an actor’s goals at the outset of a game. When such models involve the possibility of belief change, they involve the possibility that some actors will come to hold different rankings over the options before them later in the game. If such a change is caused by new information, as opposed to a change in the choice set, then rankings changes are preference changes.
communication models. He integrates Bayes’ rule—which plays an important role in many strategic communication models—into a decision-theoretic model of information processing. Calvert uses the framework to derive conditions under which a rational actor will engage in “selective exposure”—a behavior in which people pay greater attention to speakers with known biases. This work is important because many scholars use concepts such as selective exposure as assumptions in models of bounded rationality, many of which are presented as an alternative to rational actor models. Calvert, by contrast, derives selective exposure from a rational actor model (see Gerber & Green 1998 for a different view). His work is important, in the context of preference formation research because it shows how belief change is derivable from first principles about actor objectives and clarifies the incentives facing political information seekers—incentives that will affect what information they receive and which preferences they come to hold.

Gilligan & Krehbiel (1987, 1989) were among the first political scientists to use a strategic communication model explicitly to explain an aspect of politics relevant to preference formation. Their efforts clarify how amendment rules on the floor of the US House of Representatives affect the incentives of committee specialists to seek out and provide detailed information about complex legislative proposals. They derive conditions under which closed rules, which prevent the floor from amending committee bills, are beneficial to the median floor member. This beneficial effect arises because the closed rule, which appears to reduce the power of the floor median, gives committee specialists who disagree with the floor median an incentive to find and provide information that reduces everyone’s uncertainty. The closed rule works in this way because it allows powerful members on the floor of the House to pre-commit to not overturn certain committee proposals. This pre-committal, in turn, gives the specialists a reward for providing information that makes everyone better off.

This research not only reveals why powerful members of Congress may rationally relinquish their ability to amend committee proposals, it also tells us about the effects of legislative rules and institutions on legislative preferences. It shows that a simple institutional change, such as one that prevents floor amendments to committee bills, changes the quantity and character of the information available to legislators. As a result, this work provides a stirring example of how institutional choices affect the informational foundations from which legislators’ preferences can change.

Lupia & McCubbins (1998) use models and experiments to show how political institutions affect preference formation by altering a receiver’s perceptions of a speaker’s knowledge and interests. They reveal conditions under which a speaker is more likely to reveal what he knows when a third party acts as a potential verifier or when the speaker faces a penalty for lying. This kind of research unites insights from the study of persuasion with the logic of the new institutionalism to provide precise statements about how the context in which political decision-makers interact will affect the information they receive. As a result, this kind of
research clarifies the effects of particular institutional choices on the preferences that people are likely to hold.

CONCLUSION

Political scientists are active contributors to the effort to understand why people want what they want. The way forward in this effort for political science is to integrate new knowledge about human cognition and the strategic requirements of political life into existing research programs (for parallel arguments, see Denzau & North 1994, Sniderman 2000). To better understand the formation of political preferences, we must pay closer attention to extant knowledge about basic cognitive tendencies. But this is not enough, because politics is not about individual choices made in isolation; it is collective decision-making in circumstances where individual objectives cannot be achieved simultaneously. It presupposes disagreement and necessitates the possibility of strategic action. The possibility of strategic action affects the quality of information that people receive as well as their incentives for information processing. This fact leads political actors to form preferences differently than they would in many nonpolitical settings. Because the success of many policies and political endeavors (e.g. a presidential campaign, changes in the tax code, or an affirmative action policy) depends on often untested assumptions about how preferences form, substantial public value can be derived from learning more about the unique aspects of political preference formation.

ACKNOWLEDGMENT

We thank Samuel L. Popkin for helpful advice.

Visit the Annual Reviews home page at www.AnnualReviews.org

LITERATURE CITED

Brady HE, Ansolabehere S. 1989. The nature of


Hastie R, Park B. 1986. The relationship between memory and judgment depends on whether the judgment task is memory-based...
or on-line. *Psychol. Rev.* 93:258–68
Nelson TE, Oxley ZM, Clawson RA. 1997b. Toward a psychology of framing effects. Polit. Behav. 19:221–46
Reasoning and Choice: Explorations in Political Psychology. New York: Cambridge Univ. Press