The Probabilistic Mind:
Prospects for Bayesian cognitive science

Edited by

Nick Chater
University College London
UK

and

Mike Oaksford
Birkbeck College London
UK

OXFORD
UNIVERSITY PRESS

Chapter 4

Framing effects and rationality

Shlomi Sher
Princeton University, Princeton, NJ, USA

Craig R. M. McKenzie
University of California, San Diego, CA, USA

Preamble: problems of normative complexity

Most research in experimental psychology aims to construct good models of human cognitive faculties. Some important work, however, is not principally invested in the search for descriptive models of cognition. **This research aims instead to test the** empirical fit of a special class of cognitive and behavioral models, deemed interesting in their own right—the so-called 'rational actor' or **'rational choice' models (Shafir & LeBoeuf, 2002).** Even if psychologists unanimously rejected these models on empirical grounds (as most do), many would still seek to study their patterns of empirical divergence and fit. This is because of the apparent practical and philosophical significance of rational actor **models, as well as their undeniable prominence in the social** sciences.

A further strand of research, amply documented in the present volume, attempts to fuse empirical and normative studies, modeling rational actors and human actors interactively. This tradition is not interested in the classical rational actor models *per se*; it supposes, instead, that better normative models can be developed by looking closely at human behavior, and that a deeper understanding of human behavior can be achieved with the aid of suitable normative models. Some investigators in this third tradition make explicit working assumptions about the optimality of human psychological processes. Others (the present authors among them) share the basic intuition that normative and empirical analyses commonly shed valuable light on one another, albeit in sometimes subtle ways.

Still, the simple question of the empirical fit of the classical rational actor models is the central concern in several research areas in experimental psychology. This chapter examines one such area: the study of framing effects in judgment and choice.

Empirical tests of rational models run into two classes of problems. The first class is the common frustration of all experimental research in psychology: it is necessary, though often maddeningly difficult, to keep a thorough accounting of the information that is available to the subject. If the characterization of experimental inputs is incomplete, the observed outputs can severely mislead the analyst. Human cognition

seems, at first blush, a strange brew of the remarkably crude and the exquisitely subtle. For example, in studies of (explicit) visual recognition, people can be oblivious to changes even in gross details of the visual scene (Rensink *et al.*, 1997); while in studies of (implicit) visual priming, people can be highly sensitive to subtle unattended features of the visual stimulus, sometimes for weeks after a single viewing (Treisman & DeSchepper, 1996). For this reason, the ramifications of subtle information seeping unintended through an experimental design are usually difficult to prejudge. In empirical tests of all kinds—whether of rational actor models or explicitly cognitive models—the researcher must take pains to ensure that all of the information available to the subject has been accounted for.

The second class of problems is more specific to the empirical study of normative models. This research requires, not just accounting for all the information that is *available* to the subject, but also for all the information that is *relevant* to the normative model. As we will see, the latter accounting is not always easy to make.

Researchers naturally try to circumvent this second problem by examining specially contrived situations in which the normative analysis seems clear-cut—that is, in which a favored rational actor model generates clear prescriptions for normative action. One such area is framing research, an area of central importance in the psychology of judgment and decision-making. Framing researchers study situations in which apparently equivalent descriptions of choice options—for example, ground beef described as '75% lean' or '25% fat' (Levin & Gaeth, 1998)—lead to markedly divergent preferences. Normative predictions here seem particularly stark and compelling, and violations of these predictions are easy to come by. However, it turns out that, even in the simplified situations experimenters have specially contrived, the normative model used in their analysis has been inadequate. Even in this simple case, the experimental situation makes subtle information available which should matter to the normative analysis, but which has not been considered in the interpretation of experiments.

The framing literature thus affords a case study in the pitfalls of normative analysis. We do not think it is an isolated case; indeed, we will argue that closely similar problems arise in areas outside of the traditional framing literature. Such cases indicate that, in the interpretation of natural and experimental situations, adequate normative models are often as elusive as adequate empirical ones. This fact does not invalidate the empirical study of rational models, but it does highlight a basic background condition that should inform such studies. Human cognitive goals are complex. Because the function of rational norms is to guide us through our cognitive environments towards our cognitive goals, it should not surprise us if the rational norms themselves ultimately turn out to be similarly complex.

The problems and results described in this chapter illustrate this problem of normative complexity in a particularly simple empirical setting: framing effects in judgment and choice.

Framing effects: a brief review

A 'framing effect' is usually said to occur when *equivalent descriptions lead to different decisions*. Though this definition will require some amendment in what follows,

some examples will su
seek to understand:
Example 1: Evaluatir
have a terrible disease
Some participants are
while other participa
5 years'. The robust ex
treatment option when
the 'mortality' frame (
Example 2: The Asia
read the following bac

Imagine that the U.S
is expected to kill 6(
proposed. Assume th
is as follows:

Some of the participar

A: If this program is a

B: If this program is a
and a two-thirds prob

The other participants

C: If this program is a

D: If this program is a
two-thirds probability

The robust empirical
A to B, while most par
Note that, in Exam
logically equivalent, in
of the other: 20% of p
after 5 years. Similarly,
lent to C, and B is logic
Levin, Schneider, an
categories. Example 1
value a single object (he
(here, patient outcome
cally equivalent propor
valenced
shift: Preferences and
treatment options de
than options describe
beef is preferred to '25
Example 2, the
widely studied instance

some examples will suffice to illustrate the sorts of situations that framing researchers seek to understand:

Example 1: Evaluating a Medical Treatment. Participants are told to imagine that they have a terrible disease and must decide whether to accept a specific treatment option. Some participants are told that the treatment has 'a 20% mortality rate within 5 years' while other participants are told that the treatment has 'an 80% survival rate after 5 years'. The robust experimental finding is that participants are more likely to accept the treatment option when it is described in the 'survival' frame than when it is described in the 'mortality' frame (Marteau, 1989; McNeil *et al.*, 1982; Wilson *et al.*, 1987).

Example 2: The Asian Disease Problem (Tversky & Kahneman, 1981). Participants read the following background blurb:

> Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. One possible program to combat the disease has been proposed. Assume that the exact scientific estimate of the consequences of this program is as follows:

Some of the participants are then presented with the following two options:

> A: If this program is adopted, 200 people will be saved.
>
> B: If this program is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved.

The other participants instead read:

> C: If this program is adopted, 400 people will die.
>
> D: If this program is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die.

The robust empirical finding is that most participants in the first condition prefer A to B, while most participants in the second condition prefer D to C.

Note that, in Example 1, the different descriptions of the medical treatment are *logically* equivalent, in that the truth of either description necessarily entails the truth of the other: 20% of patients die within 5 years if and only if 80% of patients survive after 5 years. Similarly, an inspection of Example 2 will reveal that A is logically equivalent to C, and B is logically equivalent to D (but see Jou *et al.*, 1996; Kühberger, 1995).

Levin, Schneider, and Gaeth (1998) taxonomized framing effects into three major categories. Example 1 above is an instance of what they called *attribute framing*: the value a single object (here, a medical treatment) assumes on a single bounded dimension (here, patient outcome after 5 years) can be described in terms of either of two logically equivalent proportions (here, '% survival' or '% mortality'). When the frames are valenced (one good, one bad), the standard finding is a *valence-consistent shift*: Preferences and evaluations shift in the direction of increasing valence. Thus treatment options described in terms of 'survival' rates are rated more highly than options described in terms of logically equivalent 'mortality' rates, '75% lean' beef is preferred to '25% fat' beef, etc (Levin & Gaeth, 1988).

Example 2, the so-called 'Asian Disease Problem' is the most well-known and widely studied instance of *risky choice framing*. In framing problems in this category,

participants face two options rather than only one, and these options are gambles which can be described in terms of probabilities and proportions of gain or of loss. Usually, one option is a *sure thing* (in which an intermediate outcome is specified as certain, as in A and C above) while the other is a *gamble* (in which extreme positive and negative outcomes are both assigned non-zero probabilities, as in B and D above). The sure thing and the gamble are usually equated in expected value, making it possible to interpret observed patterns of preference in terms of participants' risk seeking or risk aversion. If we adopt this rubric of interpretation, participants encountering the Asian Disease Problem appear to be risk-averse for gains and risk-seeking for losses, a central tenet of prospect theory (Kahneman & Tversky, 1979).

Attribute and risky choice framing are widely studied[1]—the former because of its simplicity of experimental manipulation and its ubiquity in social settings (especially in persuasion situations); the latter because of its usefulness in experimentally testing classical expected utility theory and other quantitative choice models, including prospect theory.

Though risky choice framing problems are used in testing empirically oriented models, their implications for the empirical adequacy of rational actor models have been a lightning rod for debate. Attribute framing effects are of almost exclusively normative and practical interest. 'Framing effects,' Kahneman (2000, p. xv) has noted, 'are less significant for their contribution to psychology than for their importance in the real world ... and for the challenge they raise to the foundations of a rational model of decision making.'

It is important, then, to be precise about just what challenge framing effects may raise to rational actor models. According to Tversky and Kahneman (1986, p. S253), *description invariance*—the condition that equivalent descriptions must lead to identical decisions—forms '[a]n essential condition for a theory of choice that claims normative status ... so basic that it is tacitly assumed in the characterization of options rather than explicitly stated as a testable axiom'. Description invariance strikes most people as a *prima facie* reasonable normative condition, and for decades it has been generally accepted as such by psychologists.

However, there is imprecision at the heart of this formulation of the description invariance principle—*equivalent* descriptions must lead to identical decisions. What does it mean for a pair of descriptions to be 'equivalent'? While the sense of equivalence at issue is often left unspecified, the most common specification is *logical equivalence*. In this case the principle of description invariance becomes: logically equivalent descriptions must lead to identical decisions. As a theoretical criterion, logical equivalence has the virtue of transparency. Logical equivalence is well-defined (a pair of statements is logically equivalent if the truth of each entails the truth of the other); and, though disputes of application sometimes arise, they are relatively rare and, in most cases, easily circumvented.

[1] Levin *et al.*'s (1998) third category, *goal framing* (Meyerowitz & Chaiken, 1987), figures less prominently in the literature, and will not be considered here.

There is, however problem with logica

A hole in the

Every meaningful st
a cup of water on th
is 3/4-empty', 'The
strained to covary i
statements share a
forms the backdrop
respond identically t

However, the deci
kind, and a speaker
ment at random. Va
varying in degree c
speaker's selection w
speaker, as well as th
speaker's choice of f
then it also potential
rational actors would
prove relevant to the

That is, the norma
lence, but *informatic*
not only from the log
logically equivalent
be neatly separated
speakers typically sel
environments.

The condition of ir
suppose that the spea
some choice-relevant
more likely to select
not-*C*). A simple Bay
nothing special about
multiple frames.) The
ground condition *C*
probability of *C* whe
choice-relevant, we sh
fore potentially to res
When no choice-rele
exists, two frames ar
non-equivalent, and w
information (Sher &

There is, however, one problem. As we show below, there is no general normative problem with logically equivalent descriptions leading to different decisions.

A hole in the normative model: information leakage

Every meaningful statement has infinitely many logically equivalent variants. Imagine a cup of water on the table before you. The statements 'The cup is 1/4-full', 'The cup is 3/4-empty', 'The cup is 25%-full', 'The cup is twice as full as a 1/8-full cup' are constrained to covary in truth value. Though heterogeneous in style and emphasis, the statements share a common core of logical content. The normative model which forms the backdrop for all framing research insists that the decision maker must respond identically to all of these statements.

However, the decision maker must receive the statement from a speaker of some kind, and a speaker who wishes to convey this logical content will not select a statement at random. Various factors will influence the speaker's selection, these factors varying in degree of intentionality and conscious accessibility. In general, the speaker's selection will vary as a function of the information that is available to the speaker, as well as the speaker's attitudes about the thing being described. But if the speaker's choice of frame varies as a function of the speaker's beliefs and attitudes, then it also potentially conveys information about those beliefs and attitudes. Surely rational actors would not be expected to artificially ignore such information, should it prove relevant to the choice at hand.

That is, the normatively relevant equivalence between frames is not *logical equivalence*, but *information equivalence*—can any choice-relevant inferences be drawn, not only from the logical content conveyed, but also from the speaker's choice among logically equivalent statements? The normative analysis of framing effects cannot be neatly separated from the phenomena of *pragmatics*—i.e., the ways in which speakers typically select utterances and convey meaning in human conversational environments.

The condition of information equivalence can easily be formalized. For simplicity, suppose that the speaker is selecting among two frames, '*A*' and '*B*', and that there is some choice-relevant background condition *C* with the property that the speaker is more likely to select '*A*' when *C* holds than when *C* fails. That is, $P(\text{'}A\text{'}|C) > P(\text{'}A\text{'}|\text{not-}C)$. A simple Bayesian argument establishes that $P(C|\text{'}A\text{'}) > P(C|\text{'}B\text{'})$. (There is nothing special about the two-frame case: the argument immediately generalizes to multiple frames.) Therefore, a listener, aware of the regularity that relates the background condition *C* to the speaker's choice of frame, may rationally infer a higher probability of *C* when the speaker says '*A*' than when the speaker says '*B*'. If *C* is choice-relevant, we should expect a rational actor to use this information, and therefore potentially to respond differently depending on the speaker's choice of frame. When no choice-relevant background condition C meeting the above description exists, two frames are *information equivalent*. Otherwise, they are *information non-equivalent*, and we say that the speaker's choice of frame leaks choice-relevant information (Sher & McKenzie, 2006).

For framing effects to raise normative concerns, they must violate a revised principle of description invariance, which states that *information equivalent* descriptions must lead to identical decisions. The principle that logically equivalent descriptions must lead to identical decisions has no standing as a normative principle.

The formal argument that establishes the potential information content of a speaker's choice among logically equivalent frames is an elementary one. Considering the size and significance of the framing literature, this raises a natural question: why has an inadequate standard of equivalence been used so widely for so long?

There is a misleading argument that logically equivalent utterances should be treated equivalently in reasoning: if '*A*' and '*B*' are logically equivalent, there is no inference that can be drawn from knowledge that *A* that cannot be drawn from knowledge that *B*. This observation is correct, because logical implication is transitive—indeed, it is transitive with respect to forms of intuitive or probabilistic implication, which themselves may not be strictly transitive. If one knows that *A*, one knows necessarily and certainly that *B*; therefore, whatever can be inferred, logically or intuitively, from *B* can also be inferred from *A*. So *A* and *B*, when logically equivalent, must support precisely the same set of inferences, and hence, it seems, should have precisely the same effects on decision.

This normative argument would apply to framing research if it were possible for experimenters to somehow *magically endow* their subjects with knowledge that A, or alternatively with knowledge that *B*—but this is not possible. In the typical framing experiment, the participant knows that A, assuming the participant trusts the speaker, only because the participant knows that the speaker has said '*A*' . *The speaker said 'A'* is a fact which is logically equivalent neither to *A* nor to *B*. It is certainly not logically equivalent to *The speaker said 'B'* . It is true, as the above argument notes, that no inferences can be drawn from *A* which cannot be drawn from *B*. But it is false that no inferences can be drawn from the fact that the speaker said '*A*' which cannot be drawn from the fact that the speaker said '*B*' .

Because it is not possible to surgically implant statements of interest into participants' heads, the normative model we apply in experimental situations must account for the participant's knowledge that the speaker selected this statement and no other. One possible explanation for the persistence of the inadequate logical equivalence standard in framing research, then, is that experimenters may have had an idealized conception of their experimental manipulations, viewing them as implanted bits of knowledge rather than as a speaker's verbal communication (Hilton, 1995; Schwarz, 1996).

A different possible explanation for the widespread and longstanding use of a theoretically inadequate normative model in framing research is that the model is adequate for all practical purposes. Perhaps, within the linguistic domains studied by framing researchers, logical equivalence and information equivalence effectively coincide. Presumably *some* information is leaked in any speaker's selection among logically equivalent frames—but such information may be irrelevant to the listener's decision problem, or too minor to explain substantial shifts in preference. On this view, logical equivalence may not be the appropriate theoretical standard, but it is nonetheless a safe proxy standard in experimental design.

The next section wil
ment of the framing
dence for the systema
qualitatively justifies t
cally. Whether inform
in risky choice framin
section. Finally, the in
research areas falling
models employed in
which fail to take h
Information equivale
a re-examination of c
moderation of some c

Information lea

Recall that attribute
single proportion. Wl
'% lean', '% successe
'% failures'), the stan
option more highly, a
description.

But are logically eq
The parable of the ha
erwise, and experime
leak information that

To understand wh
frames, and consider
Consider, for $0 < p <$
only if *the proportion*
tosses, the proportior
portion of tosses whi

Reference Points in
Nelson, 2003; Sher &
point level of $X1$ is sa
value of $X1$. Thus, in
percentage of tosses c
hypothesis,

(1) Speakers are mc
reference point.

(2) Listeners are sen
explicitly) infer t
speaker describes

For fixed observed
descriptions when th

The next section will show that this optimistic view is untenable. In one major segment of the framing literature—attribute framing—there is strong empirical evidence for the systematic leakage of information which (a) is choice-relevant and (b) qualitatively justifies the sorts of framing effects that are commonly observed empirically. Whether information leakage is an important normative or explanatory factor in risky choice framing is less clear—this issue is tentatively explored in a subsequent section. Finally, the information leakage framework can be extended to psychological research areas falling outside of the traditional framing literature. The normative models employed in these areas, too, employ normative standards of equivalence which fail to take heed of the way information is presented to participants. Information equivalence is the needed normative standard in these areas as well, and a re-examination of classical results in the light of information leakage argues for the moderation of some classical normative verdicts.

Information leakage in attribute framing

Recall that attribute framing effects involve logically equivalent descriptions of a single proportion. When one of these descriptions has positive valence ('% survival', '% lean', '% successes') and the other has negative valence ('% mortality', '% fat', '% failures'), the standard finding is a valence-consistent shift. Participants rate the option more highly, and are more likely to select it, when it is framed with the positive description.

But are logically equivalent descriptions of proportion really information equivalent? The parable of the half-empty cup ('is the cup half-empty or half-full?') suggests otherwise, and experimental studies confirm that logically equivalent attribute frames leak information that, in typical framing experiments, is choice-relevant.

To understand what information is leaked, we need to step back from valenced frames, and consider logically equivalent descriptions of proportion more generally. Consider, for $0 < p < 1$, domains D in which *the proportion of D which is X1 is p* if and only if *the proportion of D which is X2 is* $1 - p$. For example, if D is a sequence of coin tosses, the proportion of tosses which come up heads ($X1$) is p if and only if the proportion of tosses which come up tails ($X2$) is $1 - p$.

Reference Points in Attribute Framing. The reference point hypothesis (McKenzie & Nelson, 2003; Sher & McKenzie, 2006) concerns situations in which some reference point level of $X1$ is salient to the speaker. This may be the initial, expected, or standard value of $X1$. Thus, in a sequence of coin tosses, a natural reference point value for the percentage of tosses coming up heads would be 50%. According to the reference point hypothesis,

(1) Speakers are more likely to describe D in terms of '$X1$' when $X1$ is above the reference point.

(2) Listeners are sensitive to this regularity—they are more likely to (implicitly or explicitly) infer that $X1$ is above, and $X2$ below, a salient reference point when the speaker describes D in terms of '$X1$'.

For fixed observed frequencies of $X1$ and $X2$, speakers are more likely to coin '$X1$' descriptions when the $X1$ frequency is above the reference point than when it is below

the reference point. For example, Sher and McKenzie (2006) had participants roll a six-sided die six times. For some participants, five sides of the die were black and the remaining side was white. For other participants, five sides were white and one was black. Thus, for the first group of participants, the natural reference point for the number of black outcomes out of six rolls would be five, while for the second group it would be one. Participants were not informed that the dice were weighted. For some participants, the die was weighted to fall most often on the minority color side (e.g., white in a die with five black sides). For other participants, the die was weighted to fall most often on a majority color side (e.g., black in a die with five black sides). After rolling the die six times, the participants had to describe the outcome. The reference point hypothesis predicts that, when black comes up between one and five times out of six, participants should be more likely to describe the outcome in terms of the 'black' proportion when black is the minority color on the die (and hence the black proportion is at or above reference point) than when black is the majority color on the die (and hence the black proportion is at or below reference point). This is exactly what we found. For example, when black came up three times and white came up three times, 83% of participants chose to describe the outcome as 'the die came up black three out of six times' when black was the minority color on the die, whereas only 36% did so when black was the majority color. Thus the reference point systematically influences the speaker's choice of frame.

Are listeners sensitive to such regularities in speakers' frame selection? In another series of experiments, Sher and McKenzie (2006) presented participants with two glasses of water, one full and one empty. Some participants were asked, 'just to get things started', to pour water from one glass to the other and place 'a half-full cup' in a square marked on the table. Other participants were asked instead for 'a half-empty cup.' In other experiments, different proportions were used: '1/4-full'/'3/4-empty' and '3/4-full'/'1/4-empty'.

Assuming that, for each cup, its initial state gives its reference point level, the initially empty cup, after pouring, ends up above reference point, and the initially full cup ends up below reference point. Thus the reference point hypothesis predicts that the initially full cup should be furnished more often when 'a p-empty cup' is requested than when 'a $(1-p)$-full cup' is requested. This was in fact the case, for all proportions tested. Furthermore, follow-up questionnaires revealed that the purpose of the experiment was opaque to participants—that is, they were not aware that we were testing their interpretations of proportion frames. Therefore, it is reasonable to suppose that proportion frames convey reference point information even when participants are not focused on extracting it.

These studies affirmed and extended results obtained by McKenzie and Nelson (2003) in paper-and-pencil studies. In these studies, for both cups of water and medical treatments, 'speakers' were more likely to choose an attribute frame ('% empty', '% mortality') when that attribute was above the reference point level than when it was below. Furthermore, 'listeners' were able to accurately infer the reference point from the speaker's choice of frame.

Therefore, a *rational listener* in a natural conversational environment who is uncertain about typical medical treatment outcomes *will assign* a higher probability to a

treatment's mortality rate being atypically high when the medical treatment is described in terms of its 'mortality rate' than when it is described in terms of the corresponding 'survival rate'. In general, the choice of a negatively valenced attribute frame is in fact evidence that the negative attribute is present to a greater extent than is typical. In other words, when there is uncertainty about the reference level of a choice-relevant variable, *rational actors will exhibit a valence-consistent shift.*

Implicit Recommendations in Attribute Framing. The reference point hypothesis implies that choice-relevant information is leaked in most attribute framing experiments, and that rational actors participating in such experiments would exhibit a valence-consistent shift. The empirical evidence summarized above strongly indicates that the reference point hypothesis is true. But why is it true? That is, why would speakers tend to describe objects in terms of relatively abundant attributes?

It seems likely that the speaker's choice of frame is a function, more broadly, of psycholinguistic salience, and the reference point hypothesis holds because relative abundance is one determinant of salience: more abundant attributes tend to be more salient in the speaker's psycholinguistic representations of the thing being described. But this perspective suggests that the reference point hypothesis may be profitably generalized. Reverting to the above formal terminology:

(1) A speaker is more likely to describe D in terms of '$X1$' when $X1$ is salient in the speaker's psycholinguistic representation of D.

(2) Listeners are sensitive to this regularity—they are more likely to (implicitly or explicitly) infer that $X1$ is salient in the speaker's representation of D when the speaker describes D in terms of '$X1$'.

To be sure, this generalized hypothesis is not an entirely satisfying one, because in addition to being more general than the reference point hypothesis, it is also vaguer. What psycholinguistic representations are at issue? What does it mean for an attribute to be salient in these representations? And what are the determinants of salience? However, though incomplete as a substantive hypothesis, (1)–(2) provide a valuable compass for research on information leakage in attribute framing. They set out two paths that researchers can profitably follow.

First, absent a general theory of psycholinguistic salience, one can still plausibly identify well-defined variables which are likely, all else being equal, to monotonically affect salience. Once such a variable V is identified, a corresponding pair of well-defined information leakage hypotheses is generated:

(1) Speakers are more likely to describe D in terms of '$X1$' when V assumes a higher value for $X1$.

(2) Listeners are sensitive to this regularity—when a speaker describes D in terms of '$X1$', they are more likely to (implicitly or explicitly) infer that V assumes a high value for $X1$.

The reference point hypothesis is the special case of (1)–(2) in which the variable V is *relative abundance*, one plausible determinant of salience. But many other determinants of salience could be substituted for V. For example, attributes which are more representative of the thing being described, more intrinsically notable, or more

pragmatically consequential are likely to be more salient in the speaker's psycholinguistic representations. For example, a football team with unusually dramatic and interesting victories and run-of-the-mill losses is, we suspect, more likely to be described in terms of its win rate than a team with ordinary victories and spectacular defeats. A range of such variables could be specified and the corresponding information leakage hypotheses tested. If there are exceptions to the rule—if some plausible determinants of salience fail to influence frame selection—these may particularly reward further investigation. Insofar as the rule holds up well, frame selection probability in suitably designed experiments could be employed as a measure of psycholinguistic salience in relevant areas outside of traditional framing research.

A second strategy is to employ strong and clear salience manipulations whose interpretation does not hinge on subtleties of psycholinguistic theory. Sher and McKenzie (2006) adopted this strategy in an experiment in which participants described the accomplishments of a research and development (R&D) team in a hypothetical high-tech firm. This study was the frame selection complement of a specific framing effect from the literature (Duchon *et al.*, 1989), in which an R&D team was evaluated more favorably when described in terms of its 'successful' project rate than when described in terms of its corresponding 'unsuccessful' project rate. In our experiment, participants first read a background blurb about an R&D team. For half of the participants, the blurb described an extremely impressive R&D team: the researchers were leaders in their fields, the team took on very difficult projects, the successes were revolutionary and the failures valiant, and the team was widely admired in the research community. The other participants read a blurb describing an utterly incompetent R&D team. The success/failure rate was the same for both teams, and there was no clear reference point manipulation (because the impressive team was highly skilled but also took on highly challenging projects). Participants then described the team to a hypothetical supervisor by circling words and filling in blanks. One of the three incomplete sentences forced participants to describe the team in terms of its 'success' rate or its logically equivalent 'failure' rate. As predicted, participants were much more likely to describe the impressive team than the terrible team in terms of its 'success' rate. In this experiment, there were no clear reference points, but various other factors conspired to make the successes more salient in participants' likely representations of the impressive team. This team's successes were more spectacular, more noted by the R&D community, and more representative of the team's overall high caliber.

In this way, the speaker's choice of valenced frame conveys a kind of *implicit recommendation* to the listener. That is, a rational listener can infer, from the speaker's selection of a positively valenced frame, that the favorable attribute is more likely to be salient in the speaker's representation of the thing being described—whether because the favorable attribute is relatively abundant, more representative, or otherwise notable. Hence attribute frames leak choice-relevant information about attribute salience, and this information, absorbed by listeners, justifies the ubiquitous valence-consistent shift.

Information leakage in risky choice framing

Information equivalence, rather than logical equivalence, is the needed normative standard in the analysis both of attribute and of risky choice framing experiments.

The foregoing evidence suggests that the frames studied in typical attribute framing experiments are not information equivalent, and that the leaked information is sufficient to justify the qualitative patterns of shifting preferences observed in those studies. Does information leakage have similarly strong implications for the normative analysis of risky choice effects? (Readers new to the framing literature may wish to revisit the description of the Asian Disease Problem—Example 2 above—as it will be frequently referenced in what follows.)

Risky choice framing experiments present certain challenges to an information leakage analysis. First, the descriptions communicated to participants have more moving parts: there are two options rather than just one, and various probabilities and proportions are framed conjointly. Second, the models (e.g., prospect theory) which are put forward to explain these effects are also advanced as explanations for other phenomena. Because of the relative complexity of risky choice framing problems, information leakage predictions are harder to derive. Because of the logical relationships of risky choice framing problems to other models and phenomena, these models and phenomena must ultimately be considered in any serious analysis of these problems.

However, despite their architectural and theoretical complexity, there is some reason to hope that an information leakage analysis might shed some light on risky choice framing. After all, every risky choice framing problem can be viewed as a patchwork of attribute frames (i.e., of proportions and probabilities framed in terms of gains and losses). More generally, robust implicit recommendations like those considered above may not be specific to the simple framing situations considered above.

Is there important information leakage in risky choice framing problems? The short answer is: we do not know. The available evidence is too sparse and fragmentary to undertake a serious analysis at this point. We could not hope, in particular, to do justice to the full range of evidence which argues for a prospect-theoretic interpretation of these effects. We were, however, curious to see how far the explanatory constructs developed for the simple attribute case can be extended to the more complex setting of risky choice framing. The preliminary experiments presented here are offered, not as conclusive evidence, but in the hopes of stimulating further research in this direction.

A first question is whether speakers with preferences or persuasive goals can effectively anticipate listeners' likely reactions to the different frames. That is, do speakers select frames for risky choice problems which make listeners more likely to choose the speaker's own preferred option? We have conducted a number of unpublished experiments to begin to address this question. In all of them, participants read about the sure thing and gamble in the Asian Disease Problem, but the options are *fully described* rather than selectively framed (e.g., 'If Program A is adopted, 200 people will be saved and 400 people will die'). In one experiment, participants were explicitly assigned a persuasive goal—e.g., some were to persuade a listener to choose the sure thing—and asked to pick a joint 'saved' or 'die' framing for both programs. Consistent with an information leakage account, participants indeed were more likely to pick the 'saved' framing when persuading the listener to select the sure thing than when persuading the listener to select the gamble. However, this heavy-handed manipulation

may be too artificial—e.g., speakers may deliberately simulate listener reactions to the different frames in this contrived experimental setting, but select frames quite differently in natural conversational environments. We want to understand information leakage in those environments.

To better simulate a natural conversational environment, participants in another experiment were not assigned a preference or a persuasive goal. Instead, they read the fully described options (i.e., not framed) of the Asian Disease Problem, and then indicated which program they preferred and rated their strength of preference. In this way, we were able to determine participants' personal preferences independent of framing. The participants were then presented with the following task:

> Imagine that your job is to describe the situation, and the programs which have been proposed, to a committee who will then decide which program, A or B, to use. Please complete the sentences below as if you were describing the programs to the committee.
>
> If Program A is adopted, _____ people will (write #) be saved / die (circle one)
>
> If Program B is adopted,
>
> there is _____ probability that _____ people will (write #) (write #) be saved / die (circle one)
>
> and _____ probability that _____ people will (write #) (write #) be saved / die (circle one)

Notice that, in contrast with the previous experiment, participants could independently frame the sure thing and the gamble. Regardless of prior preference (sure thing versus gamble), participants tended to frame the gamble in the same way ('1/3 probability that 600 people will be saved and 2/3 probability that 600 people will die'). However, we found a strong effect of prior preference on the framing of the sure thing. Among those who preferred the sure thing, 81% framed the sure thing in terms of lives 'saved', whereas, among those who preferred the gamble, only 48% did so. Furthermore, participants preferring the sure thing were more likely to select the 'saved' label when they rated their preference as stronger. (Those preferring the gamble were equally likely to choose the 'die' label regardless of strength of preference, indicating a possible ceiling effect for 'die' descriptions.)

This result indicates that, when given full flexibility in framing the two options, the attribute framing of the sure thing leaks choice-relevant information about the speaker's preferences. However, one limitation of this result is that participants tended to give the gamble a mixed framing throughout. To fully understand what information is leaked in standard risky choice framing problems—in which the sure thing and

gamble are jointly framed—it will be important to examine reasonably naturalistic situations in which participants nonetheless tend to choose a pure joint framing for both the sure thing and the gamble. This is because the information leaked in joint framing need not be a simple additive sum of the information leaked in the separate framing of each option.

Though preliminary, these results suggest that a deeper investigation of information leakage in risky choice framing may prove fruitful. When assigned an explicit persuasive goal, participants select the frame that would be most effective in persuading the listener. In a less constrained setting, the framing of the sure thing, at least, can leak choice-relevant information about the speaker's spontaneous preference. Further work will be needed to fully characterize the information that may be leaked in speakers' selection of joint frames in natural conversational environments.

Framing and information equivalence: new directions

Framing experiments are typically designed to scrutinize the coherence of human beliefs and decisions. However, in the analysis of these experiments, framing researchers have not sufficiently considered the relation between evidence and belief. They have generally viewed framing manipulations as implanted bits of knowledge rather than as informative utterances issued in a communicative situation. A less idealized conception of the experimental manipulations requires that we adopt a subtler normative model, subjecting frames to the standard of information equivalence rather than logical equivalence. At least in the case of attribute framing, factoring in the relation between evidence and belief undermines otherwise compelling conclusions about the coherence of beliefs and preferences. However, this problematic idealization is not unique to normative models of choice, and its problems are not unique to the experimental study of traditional framing effects.

Consider, for example, the experimental literature on hypothesis testing. In hypothesis-testing tasks, participants are commonly asked to test hypotheses of the form, 'If $X1$, then $Y1$,' where variables X and Y each have two levels ($X1$ and $X2$, $Y1$ and $Y2$). A robust finding is that participants consider an $X1\&Y1$ observation to be more supportive than an $X2\&Y2$ observation, even though both observations support the hypothesis. In other words, confirming observations that are mentioned in the hypothesis are deemed more informative than confirming observations that are not mentioned in the hypothesis (Klayman, 1995; Klayman & Ha, 1987; McKenzie, 2004b; Nickerson, 1998; Oaksford & Chater, 1994).

This tendency to consider mentioned observations maximally informative can lead to a framing effect in hypothesis testing. McKenzie and Mikkelsen (2000) had participants imagine that they were researchers investigating a possible relation between genetics and personality type. They were told that everyone has either genotype A or genotype B, and either personality type X or personality type Y. Some participants tested the following hypothesis: 'if a person has personality type Y, then he/she has genotype B' (i.e. '$Y \rightarrow B$'). Of the first two people observed, one had genotype A and personality type X ($A\&X$) and one had genotype B and personality type Y ($B\&Y$). Both observations support the hypothesis, but when asked which provided stronger

support, most participants selected the mentioned *B&Y* observation. Other participants were asked to test the hypothesis 'If a person has genotype *A*, then he/she has personality type *X*' (i.e., '$A \rightarrow X$'). Most of these participants selected the mentioned *A&X* observation as most supportive.

Note that the two hypotheses are logically equivalent (one is the contrapositive of the other), and therefore whichever observation supports one hypothesis most strongly must also support the other hypothesis most strongly. Nonetheless, participants selected different observations as most supportive depending on which logically equivalent hypothesis was presented to them. In other words, the framing of the hypothesis impacts the evaluation of evidence.

But are the logically equivalent framings of the hypothesis information equivalent? If, when testing $X1 \rightarrow Y1$, participants assume that $X1$ and $Y1$ (the mentioned events) are rare relative to $X2$ and $Y2$ (the unmentioned events), then the two logically equivalent framings are *not* information equivalent. And indeed, there is evidence that people do phrase conditional hypotheses in terms of rare events (McKenzie *et al.*, 2001).

If mentioned events tend to be rare, then, from a Bayesian perspective, the mentioned observation would be *normatively* more informative than the unmentioned observation. To see this, imagine testing the hypothesis that dwarfism leads to polydactyly (having more than 10 fingers). Because most people are not dwarfs and most people do not have more than 10 fingers, it would not be unusual to observe a 10-fingered non-dwarf regardless of whether dwarfism and polydactyly are related. However, although observing an 11-fingered dwarf would be unusual even if the two variables were related, it would be *very* unusual to observe such a person if there were no relation. Thus, observing an 11-fingered dwarf provides stronger support for the hypothesis than observing a 10-fingered non-dwarf, because the former is rare and the latter is common. (For formal details on why rarity matters, see McKenzie & Amin, 2002; McKenzie & Mikkelsen, 2000, 2007; McKenzie, 2004a; see also Anderson, 1990; Oaksford & Chater, 1994.)

In other words, treating mentioned observations as most informative is normatively justifiable because hypotheses tend to be phrased in terms of rare events. This provides a rational explanation of the fact that 'listeners' consider different data most supportive when hypotheses are rephrased in logically equivalent ways: the speakers' phrasing of a conditional hypothesis leaks normatively relevant information about event rarity.

Furthermore, the framing effect is reduced when it is clear to participants which events are rare. When this is the case, participants no longer need to rely on how hypotheses are phrased to infer event rarity. For example, the framing effect was reduced when participants were told that few people have a particular personality type and genotype. The reduction of the framing effect was especially marked when concrete hypotheses (regarding psychosis and being HIV+) were used, allowing participants to tap into real-world knowledge about rarity. Indeed, when participants were presented with concrete hypotheses and 'reminded' which events were rare, the framing effect virtually disappeared: participants were likely to select the rare observation as most informative regardless of whether it was mentioned in the hypothesis.

Thus, the application of the information equivalence standard to the normative analysis of conditional hypothesis testing helps us to understand why framing effects occur in hypothesis testing—the phrasing of conditional hypotheses leaks normatively relevant information about event rarity—and why they disappear—when event rarity is known, listeners no longer need to infer event rarity from the phrasing of the hypothesis.

McKenzie and Mikkelsen (2007) have recently made similar arguments in a discussion of human covariation assessment (for reviews, see Allan, 1993; McKenzie, 1994). In covariation assessment tasks, participants judge the strength of the relationship between two variables, each of which assumes values of presence and absence. Because people tend to give the most weight to joint presence observations, logically equivalent presentations of data can lead participants to report different judgments of covariation—a framing effect. However, because the presence of named variables tends to be rare and their absence common—e.g., there are fewer red things than non-red things, fewer accountants than non-accountants—joint presence is normatively more informative than joint absence from a Bayesian point of view. Furthermore, as with hypothesis testing, framing effects in covariation assessment virtually disappear when participants know which events are rare and which are common.

The standard normative models of covariation assessment and hypothesis testing consider only the logical content of a conditional hypothesis or data array. While these models may be well-suited to an analysis of the manipulation of idealized bits of knowledge, they are not adequate to an analysis of judgments based on specific utterances received in a complex linguistic environment. For typical abstract covariation assessment and hypothesis testing scenarios, these utterances often turn out to be information non-equivalent, leaking information about event rarity which qualitatively supports the positive conjunction strategies which are commonly observed. Determining just how far the information leakage account can go in explaining detailed results from the covariation assessment and hypothesis testing literatures would require a fuller treatment than we can provide here. However, in evaluating any given experiment in those literatures, the question of the existence and significance of information leakage from data formatting or conditional phrasing should be addressed.

Other ostensibly counter-normative phenomena, in which data or options which seem logically equivalent are treated non-equivalently, may benefit from a similar approach. McKenzie *et al.* (2006) took this approach to default effects in public policy, in which an alternative, often of considerable practical or moral significance, is more likely to be selected when it is designated as the default option—i.e., the option that will take effect barring an explicit decision to the contrary. For example, nations in which organ donation is the legal default have much higher rates of organ donation than nations in which citizens have to explicitly declare themselves donors (Johnson & Goldstein, 2003). Such 'default effects' are theoretically interesting because the options are the same in either case (e.g., 'Should I be an organ donor or not?'). While various interpretations of this phenomenon are available, information leakage may be one significant factor: the default option may serve as a kind of implicit recommendation from the policy maker to the individual, an implicit endorsement of a course of action. McKenzie *et al.* (2006) presented evidence that people draw such inferences

from the designated default. For example, they found that participants were more likely to infer that the policy makers probably thought that people ought to be organ donors when being an organ donor was the default compared to when not being an organ donor was the default. The authors also found that participants view the default to either be enrolled or not enrolled in a retirement plan as implicit financial advice. Because our normative models commonly abstract away from the way in which information is presented to decision makers—and hence they abstract from potentially important information that may leak through the speaker's choice of presentation mode—it is plausible that further examples of seemingly counter-normative behavior shine light on deficiencies, not in our everyday decisions, but in the simple normative models we use to evaluate them.

Conclusion

Experimentalists continually worry about information leaks in their research designs—they need to understand exactly what information is available to the participant if they are to understand how the participant makes use of this information. But experimentalists testing the empirical fit of normative models must worry about another kind of information leakage—they must specify exactly what subset of the available information is relevant to the proper normative model of the experimental situation. We have argued that, in the traditional framing literature, as well as in the literatures on covariation assessment, hypothesis testing, and default effects, researchers have employed normative models which are insufficiently sensitive to subtle information leaked in experimental environments. Some important results from these areas are qualitatively consistent with the hypothesis that participants are simply more sensitive to this leaked information than the idealized normative models which researchers use to evaluate their behavior.

This paper has focused on the *complexity of information* available even in simple experimental situations. Another important factor, not considered here, is the *complexity of human cognitive goals*. For instance, in normative analyses of our epistemic interactions with the world, consistency is often regarded as an end in itself. Even consistency, however, should ultimately be viewed as a means to a more sophisticated epistemic end. This is made plain in the so-called 'preface paradox': I reasonably believe that some of my beliefs are false, even though this belief renders my total class of beliefs inconsistent. Logical consistency of beliefs is a simple and compelling cognitive norm, highly useful if applied locally and with normal discretion, but it is ultimately too simple. Even if our only cognitive goals are goals of understanding, both the complexity of goals and the complexity of information situations raise formidable hurdles to the formulation of prescriptively adequate normative models.

These problems, while formidable, should not deter researchers from critically examining the rationality of human thought and action. Many phenomena of pressing social importance seem difficult to understand without some kind of irrationality assumption, and these phenomena are too significant not to try to understand. Nonetheless, in attempting to study irrationality phenomena with the aid of simple normative models in contrived experimental situations, we should proceed with caution.

Human com
obstinate con
to such subtl

Reference

Allan, L. G. (15
Bulletin, 11
Anderson, J. R.
Duchon, D., D
the facts ar
Hilton, D. J. (1
judgment.
Johnson, E. J., &
Johnson-Laird,
an introdu
Jou, J., Shantea
the role of
Kahneman, D.
frames. Car
Kahneman, D.,
Econometri
Klayman, J. (15
32, 385–41
Klayman, J., &
testing. *Psy*
Kühberger, A. (
Behavior ar
Kühberger, A. (
Organizatic
Levin, I. P., & G
informatior
15, 374–37
Levin, I. P., Sch
and critical
Processes, 7
Marteau, T. M.
patients. *Br*
McKenzie, C. R
assessment
McKenzie, C. R
defensible.
McKenzie, C. R
(Eds.), *Blac*
McKenzie, C. R
than right c

Human communicative situations are commonly awash in subtle cues. Despite our obstinate confusions and our crudeness of understanding, we are often more sensitive to such subtle information than we realize.

References

Allan, L. G. (1993). Human contingency judgments: rule based or associative? *Psychological Bulletin*, 114, 325–448.

Anderson, J. R. (1990). *The adaptive character of tought*. Hillsdale, NJ: Erlbaum.

Duchon, D., Dunegan, K. J., & Barton, S. L. (1989). Framing the problem and making decisions: the facts are not enough. *IEEE Transactions on Engineering Management*, February, 25–27.

Hilton, D. J. (1995). The social context of reasoning: conversational inference and rational judgment. *Psychological Bulletin*, 118, 248–271.

Johnson, E. J., & Goldstein, D. (2003). Do defaults save lives? *Science*, 302, 1338–1339.

Johnson-Laird, P. N., & Shafir, E. (1993). The interaction between reasoning and decision making: an introduction. *Cognition*, 49, 2, 1–9.

Jou, J., Shanteau, J., & Harris, R. J. (1996). An information processing view of framing effects: the role of causal schemas in decision making. *Memory & Cognition*, 24, 1–15.

Kahneman, D. (2000). Preface. In D. Kahneman, & A. Tversky (Eds.), *Choices, values, and frames*. Cambridge: Cambridge University Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, 47, 263–291.

Klayman, J. (1995). Varieties of confirmation bias. *Psychology of Learning and Motivation*, 32, 385–418.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94, 211–228.

Kühberger, A. (1995). The framing of decisions: a new look at old problems. *Organizational Behavior and Human Decision Processes*, 62, 230–240.

Kühberger, A. (1998). The influence of framing on risky decisions: a meta-analysis. *Organizational Behavior and Human Decision Processes*, 75, 23–55.

Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research*, 15, 374–378.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: a typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, 76, 149–188.

Marteau, T. M. (1989). Framing of information: its influence upon decisions of doctors and patients. *British Journal of Social Psychology*, 28, 89–94.

McKenzie, C. R. M. (1994). The accuracy of intuitive judgment strategies: Covariation assessment and Bayesian inference. *Cognitive Psychology*, 26, 209–239.

McKenzie, C. R. M. (2004a). Framing effects in inference tasks—and why they are normatively defensible. *Memory & Cognition*, 32, 874–885.

McKenzie, C. R. M. (2004b). Hypothesis testing and evaluation. In D. J. Koehler, & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 200–219). Oxford: Blackwell.

McKenzie, C. R. M., & Amin, M. B. (2002). When wrong predictions provide more support than right ones. *Psychonomic Bulletin and Review*, 9, 821–828.

McKenzie, C. R. M., Ferreira, V. S., Mikkelsen, L. A., McDermott, K. J., & Skrable, R. P. (2001). Do conditional hypotheses target rare events? *Organizational Behavior and Human Decision Processes*, **85**, 291–309.

McKenzie, C. R. M., Liersch, M. J., & Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science*, **17**, 414–420.

McKenzie, C. R. M., & Mikkelsen, L. A. (2000). The psychological side of Hempel's paradox of confirmation. *Psychonomic Bulletin and Review*, **7**, 360–366.

McKenzie, C. R. M., & Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology*, **54**, 33–61.

McKenzie, C. R. M., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: reference points, frame selection, and framing effects. *Psychonomic Bulletin and Review*, **10**, 596–602.

McNeil, B. J., Pauker, S. G., Sox, H. C., Jr., & Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *New England Journal of Medicine*, **306**, 1259–1262.

Meyerowitz, B. E., & Chaiken, S. (1987). The effect of message framing on breast self-examination attitudes, intentions, and behavior. *Journal of Personality and Social Psychology*, **52**, 500–510.

Nickerson, R. S. (1998). Confirmation bias: a ubiquitous phenomenon in many guises. *Review of General Psychology*, **2**, 175–220.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, **101**, 608–631.

Oaksford, M., & Chater, N. (2003). Optimal data selection: revision, review, and re-evaluation. *Psychonomic Bulletin and Review*, **10**, 289–319.

Rensink, R.A., O'Regan, J. K., and Clark, J. J. (1997). To see or not to see: the need for attention to perceive changes in scenes. *Psychological Science*, **8**, 368–373.

Schwarz, N. (1996). *Cognition and communication: judgmental biases, research methods, and the logic of conversation*. Mahwah, NJ: Erlbaum.

Shafir, E., & LeBoeuf, R. A. (2002). Rationality. *Annual Review of Psychology*, **53**, 491–517.

Sher, S., & McKenzie, C. R. M. (2006). Information leakage from logically equivalent frames. *Cognition*, **101**, 467–494.

Treisman, A. & DeSchepper, B. (1996). Object tokens, attention, and visual memory. In T. Inui, & J. McClelland (Eds.), *Attention and performance XVI: information integration in perception and communication* (pp. 15–46). Cambridge, MA: MIT Press.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, **211**, 453–458.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, **59**, S251–S278.

Wilson, D. K., Kaplan, R. M., & Schneidermann, L. J. (1987). Framing of decisions and selections of alternatives in health care. *Social Behaviour*, **2**, 51–59.

The con
for unde
tions, la
aspects
work: i.e
these rel
what ha
or reneg
majority
(see, for
[2000] a

The st
analysis
conditio
anteceden
licenses
(MT):

These
it can be
it is true
nose, the
MT infe