



Information leakage from logically equivalent frames [☆]

Shlomi Sher ^{*,1}, Craig R.M. McKenzie

Department of Psychology, University of California, San Diego, USA

Received 23 May 2005; revised 11 October 2005; accepted 2 November 2005

Abstract

Framing effects are said to occur when *equivalent* frames lead to different choices. However, the equivalence in question has been incompletely conceptualized. In a new normative analysis of framing effects, we complete the conceptualization by introducing the notion of *information equivalence*. Information equivalence obtains when no choice-relevant inferences can be drawn from the speaker's choice of frame. We show that, to support the normative implications traditionally attributed to framing effects, frames must be equivalent in this sense. We also present new evidence for McKenzie and Nelson's (2003) reference point hypothesis, which posits a tendency to cast descriptions in terms of what has increased relative to the reference point. This leakage of information about relative state violates information equivalence, and gives rise to a normative account of the most robust finding in the attribute framing literature—the valence-consistency of preference shifts. We argue that, more generally, valenced descriptions leak information about perceived valence. Such “implicit recommendations” may

[☆] This research was supported by National Science Foundation Grants SES-0079615 and SES-0242049, and a National Science Foundation Graduate Research Fellowship. The authors are grateful to Vic Ferreira and Norbert Schwartz for helpful comments on a previous draft. Some of these results were presented at the 40th Annual Bayesian Research Conference, Los Angeles, CA (2002); 8th Behavioral Decision Research in Management Conference, Chicago, IL (2002); 43rd Annual Meeting of the Psychonomic Society, Kansas City, MO (2002); the 19th Biannual Conference on Subjective Probability, Utility, and Decision Making, Zurich, Switzerland (2003).

* Corresponding author.

E-mail address: ssher@princeton.edu (S. Sher).

¹ Present address: Psychology Department, Princeton University, Princeton, NJ 08544, USA.

generalize the reference point explanation of the valence-consistent shift. Normative and psychological implications of the information leakage framework are discussed.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Framing effects; Pragmatics; Rationality

1. Introduction

A framing effect is said to occur when equivalent descriptions of a decision problem lead to systematically different decisions. Framing effects thus purportedly violate a bedrock principle of “description invariance”, “[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” (Tversky & Kahneman, 1986, p. S253). It is largely from the robust existence of framing effects that Tversky and Kahneman (1986) conclude that “no theory of choice can be both normatively adequate and descriptively accurate”. (p. S251).

This paper raises a basic question about the basic principle of description invariance and about the standard definition of “framing effect”. The standard characterization of framing effects refers, as above, to “equivalent descriptions of a decision problem”—but what does it mean for a pair of descriptions to be “equivalent”? And what *must* it mean for a pair of descriptions to be equivalent if equivalent descriptions leading to different decisions is to raise normative eyebrows? That is, just what is the *invariance* in “description invariance”? To our knowledge, these elementary questions have not been satisfactorily addressed in the literature on framing effects and description invariance. In what follows, we propose an explicit characterization of the normatively relevant equivalence—“information equivalence”—and present experimental results which suggest that an important segment of the framing literature has been concerned with information non-equivalent descriptions.

Much of the time, the questions posed above are not even implicitly addressed. Furthermore, when “equivalence” is explicated, the explication typically proceeds via appendage of a single, unexplicated adjective: for example, equivalence may be fleshed out as “objective equivalence” (Dunegan, 1996; Levin, Schneider, & Gaeth, 1998) or “formal equivalence” (Frisch, 1993; Kühberger, 1998). In our view, such adjectives add little more than emphasis.

A handful of researchers have been explicit about the sort of equivalence they have in mind—namely, logical equivalence (see, e.g., Johnson-Laird & Shafir, 1993; Rubinstein, 1998; Shafir, 1993). This explication is substantive: logical equivalence is well-defined (a pair of statements is logically equivalent if each member of the pair necessarily entails the other) and, provided some care is taken in translating between logical connectives and natural language connectives, straightforward to

diagnose. However, there is no general normative problem with *logically* equivalent descriptions of a choice problem leading to different choices.²

To see this, let A and B be a pair of logically equivalent statements about a choice problem. Suppose in addition that speakers' conversational behavior exhibits the following regularity: speakers, choosing between uttering "A" and uttering "B", are more likely to utter "A" when some background condition C (not explicitly specified in the statements A and B) holds than when C fails. In that case, a listener who hears a speaker say "A" can safely infer a higher probability of C being true than if the speaker had said "B" (that is, $p(C|\text{speaker says "A"}) > p(C|\text{speaker says "B"})$). If knowledge about the background condition C is relevant to the choice at hand, then the speaker's (e.g., experimenter's) utterance of the two logically equivalent statements A and B may with impunity lead to different decisions.

When there is no choice-relevant background condition C about whose probability a listener can draw inferences from the speaker's choice between frames A and B, we say that A and B are "information equivalent". Otherwise, we say that there has been information leakage from the speaker's choice of frame, and that the frames are therefore information non-equivalent. The simple argument from the previous paragraph shows that, for description invariance to carry any weight as a normative principle, the invariance in question must be *information invariance*.

It may seem paradoxical that different inferences can be drawn from different but logically equivalent frames, A and B. Indeed, there is no statement S that can be inferred from A but not from B (even if S is a statement about the probability of certain background conditions C obtaining), because, if one knows that B, one can also know that A, and hence draw whatever inferences one would draw knowing that A. However, when one encounters a frame A in a framing problem (or in a natural conversational environment), one is not merely endowed with knowledge that A. Instead, one is endowed with knowledge that the speaker said "A" (and not "B"). Information non-equivalence stems *not* from those inferences

² At least two other substantive explications of equivalence have been put forward in the literature, but neither is unproblematic. Tversky and Kahneman's (1986) paper appeals to both. (1) Omitting problematic reference to equivalent descriptions, they refer to different descriptions of the *same problem* leading to different decisions. However, taken literally, this characterization is inadequate. For example, if you have to choose whether to let the axe-murdering priest who wants to use your telephone into your house, it ought to make a difference to you whether you are asked whether you would be willing "to let this priest into your house to use the phone" or, alternatively, "to let this axe-murderer into your house to use the phone". The focal issue is the information contained in the description of the decision problem, and not unmentioned facts about the decision problem. Because virtually all choice-task descriptions underdetermine relevant aspects of the state of the described world, we cannot speak of whether two descriptions describe *the same problem*, but rather whether the set of problems which the descriptions truly describe is the same. That is, we are back to logical equivalence. (2) Alternatively, two descriptions of a decision problem are said to be "equivalent" if, on reflection, people would endorse their equivalence. However, no normative theory of decision making has anything to say about the correctness of people's beliefs about equivalence—which, as we argue here with reference to most framing researchers' beliefs, may after all be wrong.

which can be drawn from A but not from B (there are none), but instead from those inferences which can be drawn from the fact that the speaker said “A” rather than “B” (there are many).³

Information non-equivalence of logically equivalent descriptions has been demonstrated in other contexts. For example, Johnson-Laird (1968a, 1968b) argued that passive-form sentences and their logically equivalent active-form counterparts convey different information about the relative prominence of the logical subject and the logical object of the sentences (e.g., in “The man was kissed by the woman”, the man is intended and interpreted to be more prominent than in “The woman kissed the man”). Similarly, transposing the subject and object around a symmetrical action verb leaves logical content undisturbed, but nonetheless conveys information about causal agency: the subject of a symmetrical action predicate (e.g., the man in “The man danced with the woman”) is typically the causal initiator of the described action (Semin & De Poot, 1997; for a general analysis of “the asymmetrical behavior of symmetrical predicates”, see Gleitman, Gleitman, Miller, & Ostrin, 1996). That is, different perceptions (of relative prominence, causal agency, etc.) lead speakers to choose different sentence forms, and listeners are able to draw corresponding conclusions from the speaker’s choice of sentence form. In the above cases, two logically equivalent sentences are information non-equivalent, because information (about relative prominence, causal agency, etc.) “leaks out” from the speaker’s choice of sentence form. (For additional examples, see Moxey & Sanford, 2000; Wason, 1965.)

A skeptic might agree with our conceptualizations of “equivalence” and “invariance”, and concede the inadequacy of the standard characterization, but nonetheless maintain that the logically equivalent frames used in actual framing studies in the literature are in fact information equivalent—or at least that any information non-equivalence is too marginal to account for major shifts in preference. In the remainder of this paper, we attempt to satisfy the skeptic by documenting actual instances, of direct relevance to the framing literature, in which logically equivalent frames are demonstrably information non-equivalent. We begin by normatively re-examining McKenzie and Nelson’s (2003) reference point hypothesis in the context of the information leakage framework. We present new evidence (Experiments 1–4) for this hypothesis, overcoming methodological shortcomings in the original experiments, and we argue that, when frames are valenced, reference point information is widely choice-relevant. Extending the reference point hypothesis, Experiment 5 provides evidence that speakers tend to describe positively evaluated things in positive terms, even in the absence of a salient

³ Note that the present analysis makes no assumptions about the existence of Gricean norms, or, more generally, about the communication of informative intent (Sperber & Wilson, 1986). The analysis simply points out that, when a certain kind of regularity in speaking behavior exists, a particular kind of inference will typically be warranted, norms and intentions aside. Whether and how listeners, in drawing such inferences, consider informative intentions or conversational norms is a question for further research to address. (For Gricean perspectives on research in judgment and decision making, see Hilton, 1995; Schwarz, 1994.)

reference point. In this way, a speaker's choice among valenced frames may communicate a kind of implicit recommendation to the listener. Because of the broad relevance of relative quality and perceived valence, the information leakage documented here suggests a natural explanation for the most robust finding in the attribute framing literature—the valence-consistency of shifts in preference. In short, in many framing experiments, choice-relevant information does in fact leak out from the experimenter's choice among logically equivalent frames. Finally, we distinguish between information leakage as a normative analysis of framing problems and information leakage as a psychological hypothesis about why some framing effects occur, and consider our evidence from both angles.

2. How reference points influence frame selection

Reference points—the initial, expected, or standard level of a variable, in contrast with which other objects are implicitly evaluated—have been shown to influence a wide variety of behaviors, ranging from judgments of physical distance (Hirtle & Jonides, 1985; Sadalla, Burroughs, & Staplin, 1980) to speakers' use of Lakoff's (1973) linguistic “hedgies” (Rosch, 1975). Examining the effect of reference points on the linguistic behaviors which form the subject of the framing literature, McKenzie and Nelson (2003) showed that speakers' frame selection and listeners' frame interpretation are systematically influenced by implicit reference point information. While they did not directly address its normative implications, we argue that this finding undermines the assumption of description invariance in a sizeable segment of the framing literature.

Specifically, consider, for $0 \leq p \leq 1$, 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, in describing people undergoing a medical treatment (D), “ $X1$ ” may refer to those who die within five years of undergoing the treatment and “ $X2$ ” may refer to those who are still alive 5 years after undergoing the treatment; in descriptions of ground beef (D), $X1$ may be “lean” and $X2$ may be “fat”; and, in descriptions of a cup of water (D), $X1$ may be “full” and $X2$ may be “empty”. Much of the framing literature is concerned with just this variety of framing problem. Indeed, in a recent review, Levin et al. (1998) proposed a taxonomy of the framing literature into attribute framing, risky choice framing, and goal framing: the first of these categories is concerned with framing effects involving logically equivalent descriptions of a single proportion.

McKenzie and Nelson (2003) hypothesized the following regularity in linguistic behavior: (1) In describing a fixed state of proportionate affairs, speakers are more likely to describe the proportion in terms of “ $X1$ ” when $X1$ has *increased* relative to the reference point proportion (the norm, or what one would have expected) than when $X1$ has *decreased* relative to the reference point. (2) Listeners are sensitive to this regularity—that is, listeners are capable of correctly inferring the reference point proportion from the speaker's choice of proportion-frame. Reformulating McKenzie and Nelson's (2003) hypothesis in the terms of our normative framework, reference

point information is leaked from speakers' choices among logically equivalent descriptions of proportion, and listeners absorb this leaked information. In a later section, we argue that this leaked information is broadly choice-relevant in most of the attribute framing literature.⁴

For example, McKenzie and Nelson presented some subjects with the following scenario in a “speaker study”:

Imagine a 4-ounce measuring cup in front of you that is completely filled with water up to the 4-ounce line. You then leave the room briefly and come back to find that the water is now at the 2-ounce line. What is the most natural way to describe the cup now?

Other subjects encountered an otherwise identical scenario in which the cup was originally empty rather than originally full. Confirming the reference point hypothesis, subjects were more likely to describe the cup as “1/2 full” when it was previously empty (fullness having increased relative to the reference point) than when it was previously full (fullness having decreased relative to the reference point).

In the corresponding “listener study”, also reported in McKenzie and Nelson (2003), some subjects were presented with the following scenario:

Imagine that Mary was sitting at her kitchen table with a glass in front of her. She left the room briefly and came back to find that the contents of the glass had changed. When asked to describe the glass now, Mary said, “The glass is 1/2 full”. Given how Mary chose to describe the glass after its contents had changed, please choose the statement below in terms of what you think was most likely true about the glass before its contents changed.

Other subjects encountered the same scenario, except with Mary describing the cup as “1/2 empty” rather than “1/2 full”. Again, subjects were more likely to infer that the cup was previously full when it was described as “1/2 empty” than when it was described as “1/2 full”. In essence, subjects in the listener study correctly absorbed the information that subjects in the speaker study leaked. McKenzie and Nelson (2003) obtained similar results both using different proportions and in the domain of medical treatments, with some subjects describing treatments in terms of mortality

⁴ Some discussions of risky choice framing have been couched in terms of “reference points”, but it is important to distinguish these from the present account. In theoretical treatments of risky choice framing problems (like the Asian Disease Problem), frames are said to influence the psychological zero point in the decision maker's prospect-theoretic value function—this zero point is sometimes labeled “the reference point” (Kahneman & Tversky, 1979). However, the frames themselves are not taken to convey any information about the state of the world or of the speaker's mind. The present development, by contrast, is novel in two ways: First, it highlights the *information content* of frames—the speaker's choice of frame leaks information about an aspect of the world (the usual, expected, or initial level of the variable being described—“the reference point”, in our terminology) about which the speaker may have knowledge. Second, our account relates this reference point information implicit in frames to the attribute (as distinct from risky choice) framing literature.

or survival rate and other subjects making reference point judgments on the basis of such descriptions.

However, as the authors themselves noted, one potential shortcoming of the experiments described above is their artificiality. Specifically, participants in those studies may have felt compelled to hypothesize a relationship between the prior state of the cup and its current description. Because the relationship hypothesized by McKenzie and Nelson (2003) was intuitively generated, it is possible that subjects, acting as folk psychologists rather than spontaneous conversationalists, intuited the same relationship. The reference point hypothesis, however, concerns people's *actual* speaking and listening behavior, not their *beliefs* about actual speaking and listening behavior. To certify the McKenzie and Nelson (2003) experiments as reflecting regularities inherent in actual discourse, Experiments 1–3 recast those experiments in a more naturalistic setting with opaque manipulations. How can we determine whether, in ordinary conversational life, a person spontaneously thinks of “half-empty” cups as previously full? One way is to give her a completely full cup and a completely empty cup, ask her for a “half-empty” cup, and see what she does.

3. Experiment 1

3.1. Method

Participants were 99 UC San Diego (UCSD) undergraduates who received partial course credit for their participation. Four subjects did not provide useable data (three due to misunderstanding of the instructions, one due to experimenter error) and were excluded from the analysis. Data were analyzed for the remaining 95 subjects. In this and all subsequent experiments, subjects were randomly assigned to condition.

Subjects, tested individually, were seated at a desk in one of two small rooms. At the left edge of the desk, two apparently identical transparent plastic cups stood side-by-side. One was full of water, the other empty. (Order of the cups—whether the full or empty cup was closest to the subject—was varied between the rooms.) Indicating the right edge of the desk, the experimenter said to about half of the subjects in each room:

“Just to get things started, could you pour water from one cup to the other and set a half-full cup at the edge of the desk”.

The remaining subjects were asked for “a half-empty cup”. (Note the use of the indefinite article. Use of the definite article might have implied that the experimenter had a specific cup in mind.) After issuing this request, the experimenter left the room, giving the subject enough time to pour the water and set a cup at the edge of the desk. When the experimenter returned, the subject was given other tasks not involving the cups.

The bottom surface of each cup was imperceptibly marked by a manufacturer's number. The even-numbered cup was always initially full, the odd-numbered cup always initially empty. Therefore, once the subject had left the room, the experimenter could determine, by reading off the number, whether the initially full or initially empty cup had been furnished.

3.2. Results and discussion

Note that, after pouring, a subject would possess two cups with water up to the halfway point, one of which was initially full, the other initially empty. In line with McKenzie and Nelson's (2003) paper-and-pencil task findings, we predicted that subjects would be more likely to provide the initially full cup when "a half-empty cup" was requested than when "a half-full cup" was requested.

The results are shown by the two columns on the left side of Fig. 1. The prediction was borne out: 69% of subjects furnished the initially full cup when "a half-empty cup" was requested, whereas only 46% of subjects furnished the initially full cup when "a half-full" cup was requested ($p = .023$, 2-tailed Fisher's exact test).

These results are not susceptible to the demand characteristics objection outlined above. The dependent variable in this experiment was people's behavior in a conversational environment, not their beliefs about such behavior. Furthermore, both the experimental manipulation ("half-full" vs. "half-empty" requests) and the purpose of the experiment (surveying frame interpretation) were hidden in the between-subjects, "just to get things started" design. Indeed, occasional, informal debriefings exposed a uniform mystification about the purpose of the experiment (with many

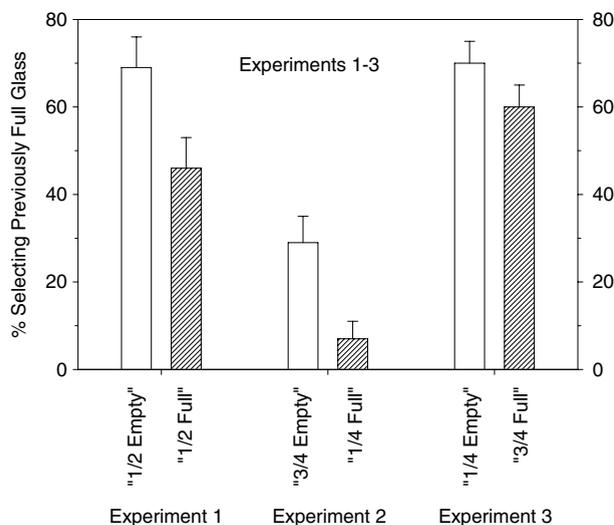


Fig. 1. Proportion of subjects selecting the initially full cup in all conditions in Experiments 1–3. Standard error bars are shown.

subjects speculating that we were interested in measurement or speed). Subjects who were queried also indicated that they felt free to furnish either cup.

Extending Experiment 1, Experiments 2 and 3 below employ different proportions, thereby advancing the naturalistic re-examination of the McKenzie and Nelson (2003) paper-and-pencil studies summarized above. They also include written rather than spoken instructions (thus ruling out the possibility of experimenter bias), and systematic checks on the opaqueness of the experimental design.

4. Experiments 2 and 3

4.1. Method

There were 112 participants in Experiment 2, and 178 participants in Experiment 3, drawn from the same population as those in Experiment 1. Five subjects were excluded from Experiment 2 and 14 were excluded from Experiment 3.⁵

As in Experiment 1, subjects were seated at a desk in one of two small rooms. Transparent cups, one full and one empty, stood side-by-side at the left edge of the desk. (Again, order of the cups was varied between rooms.)

A square, slightly larger than the base of a single cup, was marked on the desk. Subjects were given a one-page instruction sheet, which, for half of the Experiment 2 subjects, read:

In front of you on the table you're sitting at, there should be two cups and a square. To get things started, please pour water from one cup to the other and set a 3/4-empty cup in the square. Please go tell the experimenter when you've finished, and he will tell you what to do next.

The remaining subjects in Experiment 2 were asked for "a 1/4-full cup" instead of "a 3/4-empty cup".

In Experiment 3, the instruction sheet, otherwise identical to that used in Experiment 2, requested either "a 1/4-empty cup" or "a 3/4-full cup".

In both cases, after completing a series of seemingly unrelated tasks (which, for some subjects, included Experiment 5 below), subjects were given follow-up questionnaires, asking them (in Experiment 2) what they had thought the purpose of the water-pouring experiment was while participating in it, or (in Experiment 3)

⁵ Of the five unusable data points in Experiment 2, three came from subjects in the "3/4-empty" condition who gave a 3/4-full cup. Similarly, 12 of the 14 unusable data points in Experiment 3 came from subjects in the "1/4-empty" condition who placed a 1/4-full cup in the square. The high error rate presumably resulted from the fact, documented in McKenzie and Nelson (2003), that, in describing 1/4-full and (especially) 3/4-full cups, speakers have a marked preference for "full" over "empty" descriptions. Consequently, some readers in Experiments 2 and (especially) 3 probably expected to see "full" where "empty" appeared. One Experiment 3 subject was, erroneously, not administered the follow-up questionnaire.

whether they had realized that there were two ways of pouring the water to comply with the instructions, and, if so, what they thought about at the time in deciding how to pour.

4.2. Results and discussion

As in Experiment 1, the reference point hypothesis predicts that subjects will be more likely to furnish the initially full cup when “a p -empty cup” is requested than when “a $(1 - p)$ -full cup” is requested ($p = 3/4$ in Experiment 2, $1/4$ in Experiment 3).

The middle and rightmost pairs of columns in Fig. 1 depict the results of Experiments 2 and 3, respectively. In Experiment 2, 29% of subjects selected the initially full cup when “a $3/4$ -empty cup” was requested, whereas only 7% of subjects selected the initially full cup when “a $1/4$ -full cup” was requested ($p = .005$, 2-tailed Fisher’s exact test). In Experiment 3, the results were weaker (and not significant) but once again pointed in the predicted direction: 70% of subjects furnished the initially full cup when “a $1/4$ -empty cup” was requested, whereas only 60% of subjects furnished the initially full cup when “a $3/4$ -full cup” was requested ($\chi^2[1, N = 164] = 1.9, p = .17$).⁶

A glance at Fig. 1 reveals a striking difference between Experiments 2 and 3. In Experiment 2, a majority of subjects furnished the initially empty cup, while, in Experiment 3, most subjects furnished the initially full cup. This discrepancy arises from a subtle but important property of the task structure in Experiments 2 and 3. In these experiments, subjects had to *choose among two ways of pouring* (a little bit or a lot out of the initially full cup), while, after pouring, only one cup was appropriate for selection. (In this respect, they differ from Experiment 1, in which subjects could only pour in one way, and had to choose between two equivalent cups.) Because, in Experiments 2 and 3, most subjects would presumably find it easier to pour a little rather than a lot from the initially full cup, one would predict (and we found) a marked bias in favor of that pouring method. In Experiment 2, pouring a little entails selection of the initially empty cup, which, after pouring, is $1/4$ -full. In Experiment 3, pouring a little entails selection of the initially full cup, which, after pouring, is $1/4$ -empty. It is not surprising that subjects typically opted for the easier pouring method. What is interesting is how the magnitude of this preference was modulated by the experimenter’s choice of frame.

The follow-up question data in Experiments 2 and 3 confirm the opaqueness of the design. There was no indication that subjects felt compelled to intuit the reference point regularity. A closer look at the Experiment 3 responses sheds some light on underlying cognitive processes. After reading that, “[t]o comply with the instructions you were given in that experiment, you could have either poured $1/4$

⁶ It is noteworthy that McKenzie and Nelson (2003) also obtained their weakest results in the $3/4$ -full/ $1/4$ -empty conditions of both their speaker and listener studies, suggesting that in such descriptions the speaker’s choice of frame may be least informative. In both of their studies, their results for this condition were in the predicted direction, but were either not at all or only marginally significant.

of the water out of the full cup and placed the initially full cup in the square or poured 3/4 of the water out of the full cup and placed the initially empty cup in the square”, subjects were asked: “*At the time* you participated in the experiment in the small room, did you realize that you could have poured the water in either of the two ways described above”? The “framing effect” in Experiment 3 stems from subjects who reported *not* having considered both pouring methods. Among subjects who reported having done so, there was no significant difference between the proportions selecting the initially full cup in the “3/4 full” and “1/4 empty” conditions (77 and 74%, respectively). Irrespective of condition, most subjects in Experiment 3 chose the more efficient pouring method when both methods were contemplated. However, among subjects who reported *not* having considered both methods, 59% (13 of 22) furnished the initially full cup when a “1/4-empty” cup was requested, while only 16% (4 of 25) furnished the initially full cup when a “3/4-full” cup was requested ($p = .002$, 2-tailed Fisher’s exact test). Apparently, subjects who automatically “saw” a single pouring method exhibited the reference point regularity, while those who deliberated over two possibilities were guided by deliberate (e.g., efficiency) considerations. (This surmise may not generalize to Experiment 1, since, after pouring, subjects in that experiment possessed two obviously suitable cups, and it is not clear what, if any, efficiency motives would apply.) Since subjects who only considered one pouring method could not have felt compelled to intuit a relationship between pouring method and the prior state of the cup, demand characteristics could not have driven the results of Experiment 3.

In summary, in Experiments 1–3, subjects encountered one of two logically equivalent requests for a cup. Two methods of compliance differed with respect to one background condition—the prior state of the furnished cup—not explicitly specified in either frame. The reference point hypothesis states that the experimenter’s choice of frame conveys information about this background condition. Subjects reacted accordingly, providing the initially empty cup more often when a “ p -full” cup was requested than when a “ $(1 - p)$ -empty” cup was requested.

5. Experiment 4

The preceding experiments were all concerned with *frame interpretation*: the subject was presented with a description and had to act accordingly. The remaining experiments in this paper look at *frame selection*. The simplest approach to studying frame selection involves presenting subjects with a pair of prefabricated frames, and explicitly instructing them to select one (Blount & Larrick, 2000; McKenzie & Nelson, 2003). However, concerns about transparency and artificiality argue against sole reliance on this methodology. First, as noted above, frame selection tasks which are too transparent may merely measure subjects’ *beliefs* about framing behavior. Second, such artificial tasks may be too far removed from normal conversational environments, where productions are created rather than chosen.

Therefore, the remaining experiments employ a paradigm of *embedded creativity*: subjects have to construct an utterance, rather than merely choose from a set of prefabricated utterances; however, this act of construction is embedded in a highly structured task environment which: (a) is in effect as restrictive as a prefabricated choice set and (b) presumably obscures the true purpose of the experiment. Experiment 4 employs this paradigm to provide another test of the reference point hypothesis, in a new setting: describing the outcome of a sequence of independent random trials.

Imagine repeatedly rolling a die, some of whose sides are black and some white. If most of the sides on the die are black, you would expect the die to land black more than half of the time. If most of the sides on the die are white, you would expect the die to land black less than half of the time. According to the reference point hypothesis, then, you should be more likely to describe a sequence of rolls in which black comes up exactly half of the time as “half black” when the die is mostly white than when it is mostly black.

5.1. Method

Participants were 243 UCSD students, who took part immediately after having participated in Experiment 1 or in another similar water-pouring experiment. Thirteen subjects completed both parts of the questionnaire incorrectly, thus providing no useable data. Of the remaining 230 subjects, data from 3 subjects in the die-rolling component and from 5 subjects in the coin-tossing component (described below) were unuseable.

Subjects remained in the small room in which they had just completed a water-pouring task, and were given a shoebox containing a penny and a six-sided die, and a questionnaire. The first page of the questionnaire contained instructions, which the experimenter usually reiterated verbally, describing the two tasks the subject was about to complete.

The first task involved one of four black-and-white colored dice. Two of the dice were painted black on five sides and painted white on one side (i.e., black was the majority color on the die). Two of the dice were white on five sides and black on one side (i.e., black was the minority color on the die). The dice were “loaded”, although subjects were not informed of this.⁷ For each coloring of the die, one die was loaded to usually fall on the minority-color side, while the other die was loaded to usually fall on a majority-color side (specifically, on one of the four majority-color sides adjacent to the minority-color side). Each subject saw and rolled only one of the dice.

⁷ Although subtly weighted die are illegal in California, “obviously” weighted die—which fall on one side with high probability, and, on close inspection, have a distinctive roll and a very subtle indentation on the weighted side—can be sold legally. In occasional debriefings, some subjects were aware that the die they had rolled was weighted, some were unaware, and some reported vague suspicions.

Subjects filled out a form which read as follows:

Please roll the die 6 times. Record the outcome by completing the sentence below. You may use the bottom half of this page as scratch paper if you wish.

| | <u>circle one</u> | <u>circle one</u> | |
|-----------------|-------------------|-------------------|----------------|
| | | 0 | |
| | | 1 | |
| The die came up | white | 2 | |
| | | 3 | out of 6 times |
| | black | 4 | |
| | | 5 | |
| | | 6 | |

Color order (“white” on top, “black” on top) and number order (ascending, descending) were varied orthogonally with respect to one another and the type of die used. Note that, for a fixed outcome of n times black/6 – n times white, the subject, in completing the questionnaire, had to effectively choose between two logically equivalent descriptions of the outcome: “The die came up black n out of 6 times” or “The die came up white 6 – n out of 6 times”. Because of the die weighting, we could therefore examine subjects’ frame selection in describing sequences in which the minority-color side came up rarely as well as sequences in which the minority-color side came up often.

Assuming that the reference point proportion for the minority color on the die is 1/6 and that the reference point proportion for the majority color is 5/6, the reference point hypothesis makes a non-obvious prediction about how subjects will frame their descriptions. Specifically, for a fixed outcome of n times black/6 – n times white, where $1 \leq n \leq 5$, subjects should be more likely to use the “ n times black” description when black is the minority color (in which case the actual proportion for black is higher than the reference point proportion for black) than when black is the majority color (in which case the actual proportion for black is lower than the reference point proportion for black).

The next part of the experiment required subjects to flip a penny seven times. Subjects described the sequence by filling out a form much like the one above, in terms of either how many heads or how many tails came up in 7 flips. Again, the subject had to effectively choose between two logically equivalent descriptions of the sequence: “The coin came up heads n out of 7 times” or “The coin came up tails 7 – n out of 7 times”.

As presented above, the reference point hypothesis does not, strictly speaking, make predictions about the coin-flipping component of the experiment, because the reference point is fixed at .5 in all cases. However, assuming that there is a basic symmetry between “heads” and “tails” descriptions (i.e., there is never a

reference-point-independent bias in favor of one kind of description or the other), the reference point hypothesis leads us to expect that subjects will tend to describe outcome sequences in terms of the majority event (whose frequency is greater than the .5 reference point frequency) rather than the minority event (whose frequency is less than the reference point frequency). That is, all else being equal, a tendency to describe things in terms of what has increased relative to the reference point should, when the reference point is parity, favor majority descriptions.

5.2. Results and discussion

We begin with the second, coin-tossing component of the experiment. As was predicted, in describing strings of coin tosses, for each possible outcome, subjects were more likely to describe outcomes in terms of the face coming up a majority of times than in terms of the face coming up a minority of times. Pooling all of the data, 76% of subjects described their sequence in terms of the majority outcome ($p < .001$, binomial test).

This finding could be explained by a general, reference-point-independent tendency to cast descriptions of proportion in terms of the majority constituent (McKenzie, Ferreira, Mikkelsen, McDermott, & Skrable, 2001). However, such generic majority preference could not explain the predicted results of the die-rolling component of the experiment.

The die-rolling results are displayed in Fig. 2. As predicted, for each black proportion between 1/6 and 5/6, subjects were more likely to describe the outcome in “black” terms when black was the minority color on the die (and therefore the black proportion was above reference point) than when black was the majority color on the die (and therefore the black proportion was below reference point). For example,

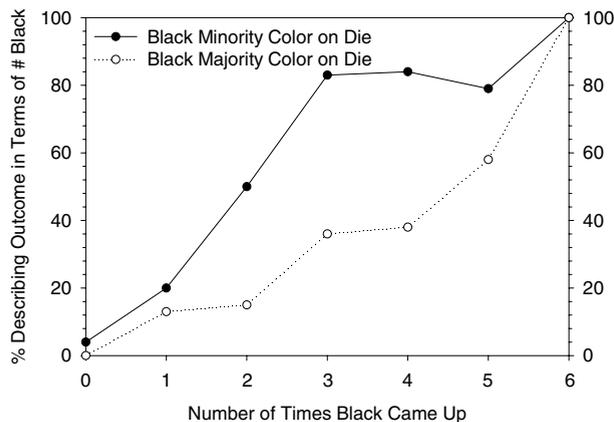


Fig. 2. For each possible sequence, proportion of subjects describing the sequence in terms of black when black was the minority color on the die (solid line) and when black was the majority color on the die (dashed line).

consider those sequences in which black and white each came up three times out of six (three on the x -axis in Fig. 2). When black was the majority color on the die, the black proportion was below reference point (because $3/6 < 5/6$), and only 36% of subjects described the sequence by saying, “The die came up black 3 out of 6 times”. However, when black was the minority color on the die, the black proportion was above reference point (because $3/6 > 1/6$), and 83% of subjects described the sequence in terms of black outcomes.

While, for each $1 \leq n \leq 5$, the “ n times black” description was chosen more often when black was the minority color than when black was the majority color, in only two cases – $n = 3$ and 4—was the comparison significant ($p < .05$, 2-tailed Fisher’s exact tests). Aggregating the data from all heterogeneous sequences, 60% of subjects used “black” descriptions when black was the minority color on the die, while only 31% of subjects used “black” descriptions when black was the majority color on the die ($\chi^2[1, N = 145] = 12.3, p < .001$). These results provide additional support for the reference point hypothesis, again with opaque manipulations. In choosing between logically equivalent descriptions of a sequence of independent random trials, subjects tend to couch descriptions in terms of what has increased relative to the reference point. Therefore, subjects’ frame selection leaks reference point information, and the frames, while logically equivalent, are not information equivalent.

6. Reference points, implicit recommendations, and the valence-consistent shift

Experiments 1–4 above provide new experimental support for McKenzie and Nelson’s (2003) reference point hypothesis, which implies that reference point information is leaked by speakers and absorbed by listeners in descriptions of frequency, proportion, and probability—the basic ingredients of attribute framing experiments. However, to establish the information non-equivalence of a pair of frames, more is required than demonstrating that information is leaked in frame selection. What is needed is a demonstration that *choice-relevant* information is leaked. The following questions thus arise: Is reference point information choice-relevant in the typical attribute framing experiment? And, if so, is observed choice behavior consistent with the rational use of leaked reference point information?

As Levin et al. (1998) observe, the most ubiquitous phenomenon in attribute framing is the “valence-consistent shift”. In the above terminology, if the statements $X1$ has proportion p and $X2$ has proportion $1 - p$ are logically equivalent, and if $X1$ has positive valence (e.g., “survival”, “success”) and $X2$ has negative valence (e.g., “mortality”, “failure”), then saying “ $X1$ has proportion p ” leads to more favorable evaluations and more sympathetic choices than saying “ $X2$ has proportion $1 - p$ ”. For example, Levin and Gaeth (1988) found that beef described as “75% lean” was more favorably evaluated than beef described as “25% fat”. Evaluations and choices shift in the direction of increasing valence.

Levin et al. (1998), extending previous work (Levin, 1987; Levin & Gaeth, 1988), advanced an associationist explanation for the valence-consistent shift: positively

valenced descriptions evoke positive associations, which in turn spawn more positive evaluations. Note that such an associationist account makes no predictions about the reference point phenomena presented above, which deal with valence-neutral descriptions and non-evaluative behaviors.

However, while the associationist explanation does not account for the reference point hypothesis predictions verified above, the reference point hypothesis suggests a straightforward explanation of the ubiquitous valence-consistent shift. Specifically, the reference point hypothesis maintains that describing a proportion in terms of “X1” signals to listeners that the proportion of X1 has increased relative to the reference point—i.e., that X1 is relatively abundant. Since it is generally good to have more of a good thing, and bad to have more of a bad thing, the reference point hypothesis predicts that proportions couched in terms of good things will lead to more favorable evaluations than proportions couched in terms of bad things (and, furthermore, the hypothesis entails that such divergence does not violate any normative principle of description invariance). That is, the reference point hypothesis indeed predicts that valenced frames will (and sometimes *should*) produce valence-consistent shifts.

To see concretely how reference point information matters when attributes are valenced, consider the evaluation of a medical treatment with a fixed survival/mortality rate. Obtaining evidence that the survival rate exceeds the norm *should* exert some upward pressure on evaluation. The reference point hypothesis predicts that the speaker’s choice of a “survival” (rather than “mortality”) frame constitutes just such evidence, and McKenzie and Nelson (2003) confirmed this prediction in their medical treatment study, described above. Replace “survival”/“mortality” with “lean”/“fat” or any other evaluative dichotomy, and the normative implication is the same. The selection of a positive frame is evidence that the positive property is more abundant than usual. Whenever considerations of relative state matter (and they usually do), a population of rational agents, cognizant of the impact of reference points on frame selection, would exhibit a valence-consistent shift.

A corollary of the reference point hypothesis, then, is that objects are more likely to be described in terms of a valenced property when that property is possessed more than some reference level. In other words, relatively positive things are more likely than relatively negative things to be described in positive terms. It is natural to ask whether this corollary generalizes to the case where there is no well-defined reference level against which comparisons are to be made. That is, are objects viewed as positive more likely to be described in positive terms than objects viewed as negative, even in the absence of any definite standard for comparative evaluation? If so, positive descriptions generally would leak a kind of implicit recommendation from the speaker: The fact that a speaker has described an object in terms of a positively valenced property makes it more likely that the speaker has a favorable attitude towards the object on the dimension under discussion.

There is good reason to expect that this corollary will generalize. Presumably, linguistic representations are typically couched in terms of those properties and events which are salient in the speaker’s psychological representation of the object being described. Hence, in the general case, the speaker’s choice of terms is a cue to

property salience in the speaker's psychological representation. In a particular conversational situation, the specific content of the leaked information will thus depend on the rules determining property salience in that situation. Viewed in this more general framework, the reference point regularity reflects the fact that relative abundance is one determinant of psychological salience: unusually abundant properties and events are more salient than unusually absent ones. However, it is unlikely that relative abundance is the only determinant of salience. In particular, properties and events which are either representative of the thing being described (e.g., the successes of a team judged to be generally successful), or intrinsically notable (e.g., the successes of a team with spectacular successes and ordinary failures), will presumably tend to be salient in psychological, and hence linguistic, representations. Since the good properties of objects deemed to be good will tend to be more representative, intrinsically notable, and abundant than the good properties of objects deemed bad, we should generally expect that psycholinguistic representations of positively evaluated objects will tend to be couched in terms of positively valenced properties and events.

Though much theoretical and empirical work would be necessary to flesh out and document this more general conception of salience in psycholinguistic representation, the specific implication that valenced descriptions leak implicit recommendations is straightforward to formulate and test. Whenever we can induce in speakers a positive evaluation of an object along some dimension, we should expect more frequent selection of positively valenced frames to describe that dimension of the object. While the reference point hypothesis predicts a valence-consistent shift resulting from information about the object's state relative to a reference point, the implicit recommendation hypothesis relaxes the reference point requirement, and hence generalizes the information leakage explanation of the valence-consistent shift. If speakers describe positive things in positive terms, listeners should infer positive things from positive terms.

Experiment 5 employs the embedded creativity paradigm to re-examine an actual attribute framing study in which a valence-consistent shift was observed. To illustrate how leaked information about perceived valence might explain this finding, we gave speaker subjects positive or negative descriptions of a team's performance and tested whether they were more likely to frame the team's performance in positive terms in the former case. This experimental design omitted any well-defined reference point, to determine whether such reference points are necessary for the leakage of information about perceived valence.

7. Experiment 5

As an example of how actual findings in the attribute framing literature can potentially be understood in terms of the communication of implicit recommendations, consider the study by [Duchon, Dunegan, and Barton \(1989\)](#), in which subjects were told to imagine that they were research-and-development supervisors at a high-tech firm. After reading a brief description of a particular R&D team, subjects had to

decide whether to raise the team's budget. As part of the description, subjects were either told that 30 of the team's last 50 projects were successful or that 20 were unsuccessful.⁸ Duchon et al. (1989) found a typical valence-consistent shift: subjects who read of the team's success rate were more generous in their funding decisions than subjects who read of the team's failure rate.

This shift in preferences poses no normative problem if speakers describe generally *successful* teams in terms of their *success* rates. More precisely, for a fixed success rate, we predict that people will be more likely to describe a successful team than a failing team in terms of its successes. We tested this prediction by presenting subjects with a caricature description of either a breathtakingly good or a catastrophically bad R&D team, and then asking them to describe the team's record to a supervisor.

7.1. Method

Subjects were 161 UCSD students, tested in groups of up to five. Subjects were given a questionnaire in an experimental session in which they had previously completed a number of tasks, including in many cases Experiment 2 or 3. Data from three subjects who completed the questionnaire inappropriately were excluded from the analyses.

Half of the subjects in the “good team” condition were presented with the following passage:

Please imagine a research-and-development project team at an international high-tech firm. This team consists of 18 people, all of whom have Ph.D.'s in Engineering. 10 of them received their Ph.D.'s at Harvard, 5 of them received their Ph.D.'s at Princeton, and the remaining 3 received their Ph.D.'s at Stanford. Three of the team members have received Nobel Prizes in chemistry and medicine for work related to their current company projects. The team generally undertakes bold and challenging problems—thus most of the team's failures have been valiant and groundbreaking attempts at seemingly unsolvable problems, while the team's successes have been truly extraordinary, sometimes revolutionary.

Of the last 50 projects undertaken by the team, 20 have been failures and 30 have been successes. The team has won the Best R&D Team Award from the International Conference of High-Tech Businesses 8 years in a row.

⁸ Note that, strictly speaking, 30/50 successful projects is not logically equivalent to 20/50 unsuccessful projects, unless one knows that every project is either successful or unsuccessful (e.g., there are no mixed or ambiguous accomplishments; one sometimes describes ventures as “not unsuccessful” to denote partial accomplishment). However, we doubt whether this wrinkle is essential to understanding the effect the original researchers found (e.g., we doubt whether carefully embedding this task in a context which made the equivalence explicit would alter the outcome), and so we will treat the frames as logically equivalent for present purposes. Others, however, have argued for the general importance of similar subtle ambiguities (Jou, Shanteau, & Harris, 1996; Kühberger, 1995).

(where, depending upon the condition, $n = 25$ or 30).⁹ We predicted that the probability of describing the team in terms of its “failures” proportion would be higher in the bad team condition than in the good team condition—that is, $p(\text{“failure”}|\text{bad team}) > p(\text{“failure”}|\text{good team})$.

7.2. Results and discussion

The experimental predictions were borne out: when the team’s record was 20 failures/30 successes, $p(\text{“failure”}|\text{bad team}) = 24\% > 3\% = p(\text{“failure”}|\text{good team})$ ($p = .007$, 2-tailed Fisher’s exact test). When the record was 25 failures/25 successes, $p(\text{“failure”}|\text{bad team}) = 34\% > 5\% = p(\text{“failure”}|\text{good team})$ ($p = .002$, 2-tailed Fisher’s exact test). That is, in choosing among logically equivalent descriptions of the team, subjects were significantly more likely to use bad terms to describe the bad team than to describe the good team.

Note that, whatever the base rates for good and bad teams, this implies that $p(\text{bad team}|\text{“failure”}) > p(\text{bad team}|\text{“success”})$. For example, with equal base rates, when the team’s record is 20 failures/30 successes, in Experiment 5 $p(\text{bad team}|\text{“failure”}) = 89\%$ whereas $p(\text{bad team}|\text{“success”}) = 44\%$. Therefore, in evaluating a message about team performance (the task in Duchon et al., 1989), a listener may be justified in inferring that a team described in terms of its “failure” rate is less promising than a team described in terms of its “success” rate.

This finding is relevant to the normative interpretation of Duchon et al.’s (1989) results, and of similar results that populate the attribute framing literature. The reference point hypothesis supplies a specific normative account for such valence-consistent shifts, because proportional descriptions couched in terms of “success” (generally, in positively valenced terms) raise the probability that the thing described is more successful (generally, more positive) than usual. The observation in Experiment 5 that, under more general conditions, positive evaluations prompt more positive descriptions suggests a more general normative account of observed shifts of preference in attribute framing experiments. The selection of a positive frame, in signaling the salience of a positive attribute in the speaker’s private representation of the thing being described, may convey a kind of implicit recommendation to the listener.

To be sure, much further work is needed to map out the conditions under which positive attitudes recruit positive descriptions. As we noted above, the notion of psycholinguistic salience is a rough one, requiring further specification; the rather

⁹ Note that, whereas Duchon et al. (1989) presented their subjects with “successful” and “unsuccessful” frames, we had our speaker-subjects chose between “successes” and “failures” frames. This is because, in a pilot study anticipating Experiment 5, in which the “successful”/“unsuccessful” formulation was used (for the 30 successes/20 failures case), few subjects chose the “unsuccessful” description *irrespective of condition* (good team or bad team). We suspected that this was because people generally avoid negations when they are unnecessary. Aware that a low ceiling on the number of subjects choosing the “unsuccessful” frame could obscure the issue of what can be inferred when a speaker does choose the “unsuccessful” frame, we avoided negations in Experiment 5 by using the “successes”/“failures” formulation (although, as is reported below, aggregating the 30 successes/20 failures data across conditions, still only few subjects chose the “failures” frame).

extreme salience manipulation in Experiment 5 was correspondingly crude. The results of Experiment 5, accordingly, leave open more fine-grained questions about the determinants of salience, and about the sensitivity of frame selection to salience. These results do, however, indicate that choice-relevant information leakage in attribute framing experiments is not narrowly confined to information about relative state. If background knowledge renders positive attributes very salient, positive descriptions become increasingly likely. Relative abundance being only one determinant of salience, the leakage of salience information may help explain why valence-consistent shifts are so large and robust.

8. General discussion

To recapitulate, we began by formulating a new normative analysis of framing effects. After showing that “framing effect” and “description invariance” have been incompletely conceptualized in the framing literature, we completed the conceptualization with an explicit characterization of *information equivalence*: a pair of frames is information equivalent if no choice-relevant inferences can be drawn from the speaker’s choice among them. There is no normative problem with logically equivalent but information non-equivalent descriptions leading to different decisions. We presented new evidence for McKenzie and Nelson’s (2003) reference point hypothesis, which posits a tendency to cast descriptions in terms of what has increased relative to a relevant reference point. This new evidence overcomes methodological limitations of McKenzie and Nelson’s (2003) work by measuring people’s behavior (rather than their beliefs about behavior) in more naturalistic conversational settings, keeping manipulations opaque, and (in Experiments 4–5) employing a frame-selection paradigm of embedded creativity. Because it is usually good to have more than usual of a good thing, information about relative state is broadly choice-relevant in the attribute framing literature. While Experiments 1–4 imply that *relatively* good things will be described in good terms, Experiment 5 suggests a generalization of this corollary: things judged to be good tend to be described in good terms. Hence, in the attribute framing literature, choice-relevant information is leaked from the experimenter’s choice of frame, and this information leakage may account for that literature’s most robust finding—valence-consistent shifts in preference. In short, subjects are reflexively sensitive to nuanced shades of meaning that are easily lost on the reflective researcher.

The information leakage framework can support two kinds of analysis of a framing finding. A normative analysis asks whether the observed behavior is, or can naturally be, justified. A psychological analysis asks about the mechanisms underlying the observed behavior. When people do the right things for the right reasons, the appropriate psychological analysis dovetails with the appropriate normative analysis. When, as sometimes happens, people do defensible things for indefensible reasons, the normative analysis has little psychological value. To be sure, matters are rarely so black-and-white: the psychological analysis may form a more or less sophisticated approximation to the normative analysis. In the framing literature, normative

analysis has traditionally played a more important role than psychological analysis. As Kahneman (2000) noted, “Framing effects 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”. However, our normative analysis—attribute framing effects are naturally justifiable in terms of sound choice-relevant inferences from the speaker’s choice of frame—suggests a psychological analysis—subjects are, at some level, actually drawing such inferences. It is worth considering each analysis in isolation.

In the present case, the normative analysis boils down to the following question: *Is the information leaked?* That is, do the frames in actual framing experiments leak enough information to plausibly justify observed preference shifts? We think that, at least with regard to much of the attribute framing literature, our findings resolve this question in the affirmative. The normative argument rests on three points: (1) Our experimental evidence, together with McKenzie and Nelson’s (2003), demonstrates that information about relative state is leaked from speakers’ attribute frame selection. (2) Information about relative state is typically *choice-relevant* when frames are evaluatively charged, because it is good to have good things in relative abundance. (McKenzie & Nelson’s (2003) medical treatment study confirms that the reference point regularity, documented here for arbitrary descriptions, generalizes to the selection and interpretation of evaluatively loaded frames, in a domain where valence-consistent attribute framing effects have been repeatedly observed (cf. Levin, Schnittjer, & Thee, 1988; Marteau, 1989; McNeil, Pauker, Sox, & Tversky, 1982; Wilson, Kaplan, & Schneidermann, 1987).) (3) We argue further that relative abundance is only one determinant of psychological salience, and present evidence that, more generally, psychologically salient properties recruit congruent linguistic terms. Together, (1) and (2) provide a self-contained normative account of the valence-consistent shift; our argument in (3), conceptually and empirically sketchier, suggests a generalization of this normative account to the communication of implicit recommendations. In either case, there is no reason to expect that a population of rational actors, drawing sound inferences from observed acts of frame selection, would behave qualitatively differently from the populations of undergraduates studied in typical attribute framing experiments.¹⁰ Attribute framing, at least, does not raise a “challenge. . .to the foundations of a rational model of decision making”.

The psychological analysis boils down to the companion question: *Is the leaked information absorbed, and, if so, how?* Experiments 1–3 point towards an affirmative answer to the absorption question, suggesting that subjects’ behavior in attribute framing experiments is driven by a general sensitivity to subtle linguistic cues. However, whether such sensitivity is appropriately conceived as *inferential* in nature remains an open, probably ill-posed question. Whatever inferences are involved

¹⁰ Our account implies that observed valence-consistent shifts are *qualitatively* normative. It neither asserts nor denies that observed shifts are *quantitatively* normative. The formulation and empirical evaluation of a quantitative normative account of information leakage in attribute framing would face significant obstacles, most notably in distinguishing between accuracy and reasonableness in frame interpretation and in measuring the information content of frames in natural environments.

are surely *implicit*—i.e., drawn below conscious awareness. Otherwise, the non-equivalence of attribute frames would have been self-evident prior to our analysis, and no disturbing conclusions about human rationality would have been drawn from attribute framing effects. Whether such implicit inferences, if they exist, can be rationally combined with explicit knowledge is another open question. If subjects *know* that a computer is randomly generating frames (cf. Schwarz, Strack, Hilton, & Naderer, 1991), will they still be sensitive to the subtle cues those frames typically convey? There is, perhaps, an analogy with sensitivity to subtle shifts in facial expression. If Bob knows that the almost imperceptible upward curls at the corners of Sue's mouth are due to a congenital nervous disorder, does this knowledge suppress Bob's impression, otherwise justified, that Sue is happy? In short, rational-seeming behavior arises from a generally functional sensitivity to subtle linguistic cues; however, the inferential malleability of this sensitivity remains undisclosed.

It is important to note that the general information leakage framework presented here is not restricted to the particular kinds of information leakage documented here. Other instances are easily recruited from our rich store of linguistic intuition. Consider, for example, the following between-subjects thought experiment:

Please imagine that you have been diagnosed with a very serious illness. This illness has a standard treatment, which leads to fatalities in a significant minority of cases. However, there is also a new treatment.

The hypothetical subject then encounters one of the following four new treatment descriptions:

- (1a) The new treatment has strong negative side effects *but* leads to 80% survival/20% mortality.
- (1b) The new treatment has strong negative side effects *and* leads to 80% survival/20% mortality.
- (2a) The new treatment has no negative side effects *but* leads to 80% survival/20% mortality.
- (2b) The new treatment has no negative side effects *and* leads to 80% survival/20% mortality.

The hypothetical subject is asked which treatment she would choose—the standard treatment or the new treatment? Presumably (1a) and (1b) would meet typical standards of equivalence (e.g., logical equivalence), if those standards are applied mechanically, because the side-effects quote and the survival rate estimate are fixed across frames; likewise for (2a) and (2b). However, we strongly suspect that (1a) will be viewed more favorably than (1b), and that (2b) will be better received than (2a). And justifiably so: the opening passage greatly underspecifies the scenario—how serious is the “serious illness”, and, more to the point, what is “a significant minority of cases”? The rhetorical choices of “but” (which suggests that the second property has valence opposite that of the first property) and “and” (which carries no such suggestion) partially resolve this ambiguity, but they resolve it differently. Since “strong negative side effects” has negative

valence, “but” in (1) suggests that the quoted survival rate has positive valence (e.g., is an improvement over the standard treatment), and therefore favors (1a) over (1b). However, since “no negative side effects” has positive valence, “but” in (2) suggests that the quoted survival rate has negative valence, and therefore favors (2b) over (2a). Indeed, we would not be surprised to find rational actors favoring (1a) to (2a) in a between-subjects reification of this thought experiment. McKenzie (2004) has extended the information leakage framework further to research areas lying beyond the traditional framing literature, arguing that logically equivalent conditional hypotheses and descriptions of correlational data can leak normatively relevant information about event rarity.

On the other hand, while information leakage explanations are not confined to reference points and implicit recommendations, they do not cover the framing literature exhaustively. For example, Kahneman and Tversky (1979) asked some subjects, told to imagine having been given \$1000, to choose between a 50% chance of gaining \$1000 and a sure gain of \$500. A large majority of subjects preferred the sure gain. Other subjects were told to imagine having been given \$2000, and asked to choose between a 50% chance of losing \$1000 and a sure loss of \$500. A large majority of subjects preferred the gamble. This despite the fact that, in both cases, the gamble leads, with equal probability, to a net gain of \$1000 or \$2000, while, in both cases, the sure thing ensures a net \$1500 gain. Though the two frames are not logically equivalent, we believe that they are, under usual conceptions of choice-relevance, information equivalent. Hence, information leakage does not account for this framing effect.

The typical laboratory setting has two properties which, we speculate, tend to exaggerate the power of information leakage in experimental findings. (1) Most experiments in judgment and decision making place the subject within an informationally impoverished environment. The less information there is, the greater the significance of new information generally, and hence of information leakage specifically. (2) In many experiments in judgment and decision making, subjects have to make a rather extraordinary sort of inference—one might call them *creative inferences*. In particular, subjects are presented with the skeleton of a scenario and their hypothetical reactions are essayed. However, because the skeleton vastly underdetermines the scenario, and hence the essayed reaction, subjects have to flesh out the skeleton with details to make the problem intelligible.¹¹ In short, the task is one of *construction* as much as of true inference, with quite often a lot to construct. Because of (1), subtle informational cues in phrasing will play a major role in the inferences the subject must draw, and, because of (2), the subject must draw many and important inferences. Therefore, it is not surprising that a little information leakage goes a long way in typical experiments in judgment and decision making.

¹¹ Indeed, fleshing out this skeleton often requires subjects to engage in a kind of role-playing, assuming an imagined identity that requires creative construction. To pick an arbitrary and not especially flagrant example (Thaler, 1999), subjects may be asked to: “Suppose you bought a case of a good 1982 Bordeaux in the futures market for \$20 a bottle. The wine now sells at auction for about \$75 a bottle. You have decided to drink a bottle. Which of the following best captures your feeling of the cost to you of drinking this bottle?”

Though our analysis questions the normative implications traditionally drawn from attribute framing effects, other implications remain undisturbed. In our view, framing effects are best understood, not as paradoxes of rationality, but as paradoxes of measurement.¹² That is, framing effects generally raise questions about whether elicitation of preferences really measures deep underlying preferences (cf. Payne, Bettman, & Johnson, 1992; Slovic, 1995). It is important to emphasize that our account does not deny that attribute framing effects have troubling implications for preference measurement. Therefore, the implications of attribute framing for the meaningfulness of public opinion surveys, contingent valuation measures, and, for that matter, elicited preferences in psychology experiments are untouched by the present development. However, our analysis and the traditional analysis locate the measurement problem in different places. Whereas framing effects *with information equivalence* raise questions about whether preferences can be said to exist at all (that is, there is nothing to measure), framing effects *without information equivalence* raise the question of whether the analysis of preferences is being undertaken at sufficiently high resolution (that is, the measurement apparatus is not sensitive enough to detect subtly leaked information that is affecting the object of measurement; Schwarz, 1999). In either case, we may safely conclude that different frames are not eliciting dependable measures of a single thing, either because, in the case of information equivalence, preferences only exist *at lower resolution than the analysis* (nothing is being measured), or because, in the case of information non-equivalence, preferences only exist *at higher resolution than the analysis* (more than one thing is being measured). In short, it is the *theoretical* (i.e., relating to the existence and rationality of preferences) and not the *practical* (i.e., relating to the reliability of preference measurement) implications of attribute framing effects that we are calling into question.

Demonstrating the information equivalence of a pair of equivalent frames will generally be much harder than demonstrating the information non-equivalence of a pair of non-equivalent frames. The former requires ruling out all possible choice-relevant inferences, while the latter requires ruling in only one. That having been said, there are times (e.g., the Kahneman & Tversky (1979) gamble described above) when it is reasonable to view a pair of frames as presumptively information equivalent. More often, however, absent a demonstration of information non-equivalence, information equivalence will at best be tenable as a provisional default

¹² Even when information equivalence is satisfied, framing effects only raise *prima facie* normative issues when we are inclined to presuppose or demand that decision makers possess completely ordered preferences. In our view, complete ordering of preferences is not a normative standard, but rather a convenient mathematical idealization which makes some decision theories formally tractable. For a decision maker with partially ordered preferences, however, the need to act will dictate a need for (possibly arbitrary) procedures for making decisions in the ambiguous no-man's-land outside of the partial ordering. Provided that the decision maker is flexible (e.g., able to adapt these arbitrary procedures in the unlikely event that a cunning salesman tries to convert behavioral intransitivity into a money pump), this incompleteness of preferences need not cause the decision maker any grief. To be sure, if preferences are partially ordered and the decision maker clings blindly to fixed ambiguity-resolving procedures, she will be vulnerable in principle to certain traps (though whether and how such vulnerability in principle gets exploited in practice is of course a separate question).

hypothesis. Conversation is a minefield of subtle contextual cues in which the framing researcher is advised to tread carefully. Everyday turns of phrase are, we suspect, typically soaked through with leaked information, perhaps in quite unexpected ways.

Appendix A

A.1. Experiment 5 Frame selection task

Please imagine that you are working in the Budget Office of the high-tech firm described on the previous page. The Budget Office needs to decide which of its R&D teams deserve increased funding and which deserve reduced funding in the next fiscal year. Accordingly, your supervisor in the Budget Office has asked you to provide some basic information about the R&D team described on the previous page, which he will consider in making funding decisions. The following are your supervisor's questions. Please provide the answers by completing the sentences in the way that seems most appropriate.

Question 1: "How educationally well-qualified are the team members?"

Your answer: The team consists of _____ researchers, of whom _____ have _____ .
 (write #) (write #) undergraduate degrees
 Ph.D.'s
 (circle one)

Question 2: "What is this team's track record?"

Your answer: Of the last 50 projects undertaken by the team, _____ were _____ .
 (write #) successes
 failures
 (circle one)

Question 3: "Has this team received any special commendations from the research community?"

Your answer: This team has _____ been commended by the research community for its contributions.
 never
 rarely
 sometimes
 often
 (circle one)

Note: Order in Question 2 ("successes" or "failures" on top) was varied orthogonally with respect to the other manipulations in Experiment 5.

References

- Blount, S., & Larrick, R. P. (2000). Framing the game: Explaining frame choice in bargaining. *Organizational Behavior and Human Decision Processes*, 81, 43–71.
- 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.
- Dunegan, K. J. (1996). Fines, frames, and images: Examining formulation effects on punishment decisions. *Organizational Behavior and Human Decision Processes*, 68, 58–67.
- Frisch, D. (1993). Reasons for framing effects. *Organizational Behavior and Human Decision Processes*, 54, 399–429.
- Gleitman, L. R., Gleitman, H., Miller, C., & Ostrin, R. (1996). Similar, and similar concepts. *Cognition*, 58, 321–376.
- Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin*, 118, 248–271.
- Hirtle, S. C., & Jonides, J. (1985). Evidence of hierarchies in cognitive maps. *Memory & Cognition*, 13, 208–217.
- Johnson-Laird, P. N. (1968a). The interpretation of the passive voice. *Quarterly Journal of Experimental Psychology*, 20, 69–73.
- Johnson-Laird, P. N. (1968b). The choice of the passive voice in a communicative task. *British Journal of Psychology*, 59, 7–15.
- 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 and 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.
- 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.
- Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, 2, 458–508.
- Levin, I. P. (1987). Associative effects of information framing. *Bulletin of the Psychonomic Society*, 25, 85–86.
- 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.
- Levin, I. P., Schnittjer, S. K., & Thee, S. L. (1988). Information framing effects in social and personal decisions. *Journal of Experimental Social Psychology*, 24, 520–529.
- 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. (2004). Framing effects in inference tasks—and why they are normatively defensible. *Memory and Cognition*, 32, 874–885.
- McKenzie, C. R. M., Ferreira, V. S., Mikkelsen, L. A., McDermott, K. J., & Skrabble, R. P. (2001). Do conditional hypotheses target rare events?. *Organizational Behavior and Human Decision Processes* 85, 291–309.
- 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.

- Moxey, L. M., & Sanford, A. J. (2000). Communicating quantities: A review of psycholinguistic evidence of how expressions determine perspectives. *Applied Cognitive Psychology*, *14*, 237–255.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, *43*, 87–131.
- Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, *7*, 532–547.
- Rubinstein, A. (1998). *Modeling bounded rationality*. Cambridge, MA: MIT Press.
- Sadalla, E. K., Burroughs, W. J., & Staplin, L. J. (1980). Reference points in spatial cognition. *Journal of Experimental Psychology: Human Learning & Memory*, *6*, 516–528.
- Schwarz, N. (1994). Judgment in a social context: Biases, shortcomings, and the logic of conversation. *Advances in Experimental Social Psychology*, *26*, 123–162.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, *54*, 93–105.
- Schwarz, N., Strack, F., Hilton, D., & Naderer, G. (1991). Base rates, representativeness, and the logic of conversation: The contextual relevance of “irrelevant” information. *Social Cognition*, *15*, 67–84.
- Semin, G. R., & De Poot, C. J. (1997). Bringing partiality to light: Question wording and choice as indicators of bias. *Social Cognition*, *15*, 91–106.
- Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory & Cognition*, *21*, 546–556.
- Slovic, P. (1995). The construction of preference. *American Psychologist*, *50*, 364–371.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Blackwell.
- Thaler, R. (1999). Mental accounting matters. *Journal of Behavioral Decision Making*, *12*, 183–206.
- Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, *59*, S251–S278.
- Wason, P. C. (1965). The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, *4*, 7–11.
- 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.