Belief Bracketing:
Can Partitioning Information Change Consumer Judgments?

Suzanne Shu
George Wu
University of Chicago
Graduate School of Business
1101 E. 58th Street
Chicago, IL 60637

Draft date: October 8, 2003

Abstract

When consumers seek out information while forming a belief about a product, they may observe information narrowly bracketed into a few observations at a time, or broadly bracketed into many observations at once. What influence does this bracketing have on their resultant judgments? We propose that when individuals form beliefs based on bracketed information, the information is first evaluated based on its relative representativeness to the hypothesis under consideration, and then integrated into the overall judgment using a process that shows strong primacy and recency effects. Since bracketing affects both the representativeness-based coding step and the integration step, the final judgment will be affected by the size of the brackets. This is tested and confirmed in a series of four studies, using a variety of measures and contexts (restaurant ratings, probabilistic judgments). Although these studies vary in the use of memory, the frequency of judgment, and the type of judgment, all four studies show belief bracketing effects. The overall finding is that judgments are more extreme for information delivered in broad brackets, as our model predicts. We also find bracketing effects for behavioral intentions and for accuracy.
When consumers make a judgment or form a belief about a new product or service, they seek out multiple sources of information. Sometimes this information is received in large segments of many observations at once, while other times the information may arrive in drips and drabs, with observations separated by temporal or even physical distance. For example, a consumer forming a judgment about a new restaurant may decide to look at reviews one or two at a time, while another may peruse a dozen all at once. Whether this partitioning of information is controlled by the firm or the consumer, is it possible that simply breaking the information pattern into smaller or larger partitions has an effect on beliefs about the product or service under consideration? This basic question motivates the research presented here.

A long history of research on how individuals integrate information into judgments has demonstrated that a consumer’s belief formation process is sensitive to a variety of influences that are not always consistent with a Bayesian inference process. While previous research has focused on source credibility, use of base rates, order of information, and biased encoding of observations, this research focuses on a different effect: with order of information, base rates, and source credibility held constant, how does the amount of information observed at each integration step influence the final judgment? Our claim is that the partitioning of information will influence the inferential process and hence the final judgment. We propose that local “bracketed” beliefs are formed at the level of each partition of information. These localized judgments depend on both content and sample size for each single partition of observations; in other words, individuals pay attention not only to the proportion of positive and negative signals, but also to the overall number of signals within that group. Rather than code the observations according to a Bayesian’s probability likelihood function, the judgment reflects a sense of how “representative” the data is of the hypothesis in question, tempered by a concern that small
samples are less informative (and thus less convincingly representative) than larger samples. The localized judgments are then integrated through an averaging model to form an overall judgment. Biases in both the coding and integration steps influence the final judgment in predictable ways. The primary influence comes through the representativeness-based coding step; because of sensitivity to sample size, overall judgments for information observed in smaller, seemingly unrepresentative partitions will be less extreme than judgments for the same information delivered in larger partitions.

To test the hypothesized process, we run a series of four studies that manipulate bracket size to see what effect bracket size has on extremity of judgment. We also look for a moderation of the overall bracketing effect based on recency and primacy influences within the patterns of observations. The first two studies use the paradigm of an individual forming judgments of restaurants based on a series of restaurant reviews. In these studies, we explicitly test for both the overall bracketing effect and the moderating effects of primacy and recency. Study 2 tests whether the bracketing effect occurs spontaneously or only when observers are asked to make interim judgments. Study 3 uses a more traditional experimental “balls and urns” paradigm to examine the influence of bracketing on probability judgments. This study is complemented by Study 4, which collects measures of representativeness for the patterns used in Study 3, and tests whether an averaging model of these representativeness measures is able to account for the extremity results in the earlier study. Overall, although these studies vary in the use of memory, how often judgments are collected, and the type of judgment, all four studies show the predicted belief bracketing effects. In particular, judgments are generally more extreme for information delivered in broad brackets. We also find bracketing effects for behavioral intentions and for accuracy.
Background

Bayes’ theorem provides a normative standard for formation and updating of consumer judgments (e.g., Hagerty and Aaker 1984). Bayes’ theorem dictates that in a sequence of independent events, the order of the events and the subsets in which they are presented should have no influence on the final judgment. All observers should produce the same posterior probability estimate as long as the sequence of observations is identical across all conditions. More formally, Bayesian updating relies on a product rule (the product of the prior odds and the likelihood ratio for all observations), which is independent of the order in which the updating occurs.

There has been considerable research to test whether individuals obey Bayesian updating in forming probability judgments, using both non-ambiguous stimuli (e.g., “balls and urns”) and attribute descriptions more susceptible to subjective interpretation. Typical findings are that individuals appear to use an averaging model to integrate information (Shanteau 1975, Anderson 1981, Birnbaum and Mellers 1983, Lopes 1985), that individuals do not give appropriate attention to base rates or the credibility of the source (Kahneman and Tversky 1973, Birnbaum and Mellers 1983), and that non-diagnostic information can weaken beliefs (Trautman and Shanteau 1976, Fischhoff and Beyth-Marom 1983). Many of these findings were later extended to product judgment contexts, including work on how consumers form preferences and/or beliefs about products and services. For example, failure to attend to source credibility has been studied in the context of using consumer agents for recommendations (Gershoff, Broniarczyk, and West 2001), and the influence of nondiagnostic information has been assessed in how consumers treat irrelevant product information (Simonson, Carmon, and O’Curry 1994, Russo, Meloy, and Medvec 1998, Chernev 2001, Meyvis and Janiszewski 2002).
While these previous efforts have informed us about how people integrate information into their beliefs, the work presented here has some notable differences. First, it differs from the normative predictions of Bayes’ theorem in suggesting that individuals do not update according to a product rule, which requires independence of the order of the information and the size of the partitions in which information is delivered. Instead, we predict that both bracketing and order will have a substantial effect on final judgment. Second, the objective and non-ambiguous information encountered by our observers differs from that used in much of the previous research. This restricts the role of motivational bias in interpreting signals; two individuals will judge a given signal similarly regardless of prior beliefs. Finally, each signal in our studies is independently and randomly generated from equally diagnostic sources, and hence equally relevant to the final judgment, unlike studies which vary source credibility. In sum, our main emphasis is not on the type of information being provided, but instead on how the partitioning of that information affects the final judgment through a process we call “belief bracketing”.

Borrowing from the definition of choice bracketing (Read, Loewenstein, and Rabin 1999), a belief bracketing effect can be defined as a situation where a series of local judgments do not match a global judgment based on the same set of information. In the examples provided throughout this paper, bracketing is operationalized as the grouping of a set of independent signals together into subsets. Large subsets are called broad brackets, while small subsets are narrow brackets. Consider, for example, a series of 50 flips of a coin. The results of all 50 flips

---

1 This concept of grouping information together into narrow or broad brackets is not new, although it is often described in different terms in other literatures. It has been described as narrow versus broad framing (Kahneman and Lovallo 1993), partitioning (Ariely and Zauberman 2000) and segregation (Langer and Weber 2002). Related concepts in information processing have been described as sequential versus simultaneous presentation of information (Kardes and Kalyanaram 1992) or piecemeal versus categorical processing (Sujan 1985). In perception research, the effects of bracketing on frequency illusions have been termed category-split effects (Fiedler and Armbruster 1994).
at one time constitute the broadest bracket, whereas 10 subsets of 5 flips each constitute a narrower bracket.

The influence of bracketing has been well documented in choice. Choice bracketing exploits the asymmetry between gains and losses (Kahneman and Tversky 1979) as well as a narrow or myopic approach to decision making (Kahneman and Lovallo 1993). Read, Loewenstein, and Rabin (1999) provide an extensive review of choice bracketing effects, such as adding-up effects (in which aggregated results are neglected) and trade-off effects (in which poor results of some choices can be integrated with the positive results of other choices). In sum, decision makers isolate current choices from a larger portfolio of choices, causing them to overweight immediate losses. For example, gambles that are presented in isolated stages are often rejected, while the same set of gambles is accepted when integrated into a single outcome. One example of choice bracketing, myopic loss aversion (Benartzi and Thaler 1995) takes advantage of both loss aversion and a lack of aggregation characteristic of narrow framing. Both Thaler, Tversky, Kahneman, and Schwartz (1997) and Gneezy and Potters (1997) test myopic loss aversion experimentally and find that narrow bracketing leads to more risk aversion (lower allocation to stocks or smaller bets). These myopic loss aversion studies theorize that the narrow bracketing results in a change in preferences rather than a change in beliefs.

Bracketing effects are less well-documented in judgment. One exception is the work of Ariely and Zauberman (2000, 2003), who found that partitioning a hedonic experience can dampen satisfaction evaluations. In these studies, bracketing an experience shifts the overall evaluation toward the experience’s mean satisfaction rating. Patterns that are frequently evaluated or otherwise narrowly bracketed receive a less extreme evaluation than patterns presented as a single experience. The effect of bracketing was most apparent for experiences
with a trend component: upward trends are judged as less satisfying when they are partitioned, while downward trends are judged as less negative when they are partitioned. Ariely and Zauberman hypothesize that this effect is the result of observers recalling only a summary satisfaction measure from each partition, rather than recalling all information at once. In other words, the observer makes a satisfaction judgment for each individual partition, and then integrates these piece-wise satisfaction judgments to form a final satisfaction measure.

Both the choice bracketing and the satisfaction partitioning effects result from a lack of information aggregation. The decision maker “misses the big picture”, failing to see how the individual pieces of information fit into a larger pattern or outcome distribution. While this failure to take a broad view of the data is also true for belief bracketing, the underlying psychological mechanisms that drive belief bracketing are quite different. In choice bracketing, the bracketed options are evaluated according to prospect theory, which relies heavily on loss aversion, and the options within one bracket are never aggregated with choices in other brackets. Belief bracketing, however, uses representativeness to evaluate the bracketed information, and then integrates new judgments using an averaging process. This integration process is similar to the judgment integration that occurs in Ariely and Zauberman’s satisfaction partitioning process. However, belief bracketing’s representativeness coding step is again the differentiating factor, in comparison to the trend-sensitive hedonic judgments that characterize the satisfaction partitioning effects.

**The Influence of Bracketing on Beliefs**

**An Example**

To motivate the theorized process underlying belief bracketing, we consider a simple example of how different levels of bracketing can affect beliefs. Imagine that an individual is
observing some sequence of binary outcomes, such as a series of coin flips or colored balls drawn from an urn (with replacement). The observer is told that the underlying process generating these observations is biased toward one of the two outcomes (e.g., a heads-biased coin, or a predominantly red urn), but she is unsure of which direction it is biased. The prior probability of the direction of bias is .5. The total sequence will consist of 10 draws of outcomes coded as 1 or 0. The sequence may be bracketed as either 2 draws at a time, 5 at a time, or 10 at a time. Figure 1 displays the same sample sequence in all 3 bracketing versions. We consider how subjects may react to these outcomes when the data are presented in different brackets.

First consider the case where the 10 observations are bracketed into subsets of 2. The first subset of 2 is ‘10’. This sequence alone seems completely undiagnostic; there is no clear evidence of bias in this draw. The second subset is ‘00’, which although biased toward 0, could easily come from either process since there are only 2 observations. The next subset is ‘10’, which, like the first subset, is again undiagnostic. The fourth subset, ‘00’, again indicates possible bias but is not convincing on its own. The final subset is again the undiagnostic ‘01’. Overall, although there is some hint of bias toward 0, the observer is not definitively convinced of it.

Now consider the case where the same set of observations is provided in brackets of size 5. The first subset is ‘10001’. Here, there are 3 0’s to 2 1’s, which, although not overly convincing, certainly suggests bias toward 0. The next subset, ‘00001’, is much more convincing, and now the observer feels pretty sure that he’s seeing a process that’s biased toward 0’s. In other words, after seeing the sequence in the bracket size of 5, a bias toward 0’s is more discernable than it was when seen in a bracket size of 2.

Finally, in the broadest bracket, where all 10 outcomes are observed at once, the sequence is
‘1000100001’. The 7 0’s, against only 3 1’s, may make the observer feel very certain that the process is biased toward 0’s. Thus, the same sequence seen in three different ways results in different final beliefs regarding the process generating these observations.

This example suggests some elements of what underlying processes might cause the bracketing effect. The primary driver is how each subset of outcomes is evaluated and then integrated with prior beliefs. We suggest that observers evaluate each subset of outcomes according to how representative it is of the underlying process (Tversky and Kahneman 1971). A bracket-level judgment is then formed based only on that particular subset. The observer uses this bracket-level judgment to update their overall judgment about the sequence as a whole. However, the order in which the brackets are observed may also have an effect, with the first and last brackets each being weighted more heavily than middle brackets.

We capture this process with a more detailed model. We start by reviewing the normative updating model, which we then use as a basis for our proposed bracketing model.

Proposed Theory

The way in which the judgments are formed and revised is normatively described by the process outlined in Bayes’ theorem. Bayesian updating comprises two distinct steps: the coding of a signal according to a probabilistic likelihood function, and the integration of that signal into the prior existing belief. As each new signal is first coded and then integrated, the resulting posterior probability becomes the prior for any new signals subsequently received. Thus, the prior always serves as a sufficient statistic for the history of what the observer has seen. It is not necessary to remember the content or order of the previous signals. It is easy to see that, with this process, the order and/or bracketing of the signals have no effect on the overall posterior probability of a hypothesis.
Research has found that individuals are roughly consistent with Bayes in how they update beliefs, but that their judgments are usually too conservative when compared to the Bayesian standard (Edwards, 1968). The cause of this conservatism is often attributed to either misperception at the coding step, misaggregation at the integration step, or both. Additional work on information integration and belief formation has delved deeper into these two steps to better understand what individuals are doing, and to better identify the sources of bias. At the coding step, it has been suggested that individuals are evaluating new information according to subjective rather than objective probability measures (Wallsten 1972), that evaluations reflect nonlinear use of response scales, that source credibility is not always properly accounted for during coding (Birnbaum and Mellers 1983), and that the interpretation of new observations is biased in the direction of currently held beliefs (Russo, Meloy, and Medvec 1998). At the integration step, the consistent experimental finding is that new evidence is often averaged into existing beliefs rather than being incorporated additively, as Bayes’ theorem would suggest (Shanteau 1975, Anderson 1981, Birnbaum and Mellers 1983, Lopes 1985). For example, while a Bayesian’s reaction to new nondiagnostic information would be to leave his judgment level unchanged, individuals in these experiments usually weaken their judgment level (Trautman and Shanteau 1976).

These findings provide a useful framework for modeling the belief bracketing process loosely described in the above example. First, there is a violation in the coding step, in which subjects use a “representativeness” function that differs from Bayesian likelihood, and does not in general obey the product rule. Second, there is a violation in the integration step, in which the updating is conservative (as Edwards found), consistent with an averaging model (as Trautman

---

2 Edwards (1968) also allowed for the possibility of conservatism being an artifact of the types of tasks his
and Shanteau and others have found), and also sensitive to order effects. Both the coding and integration violations are described here, and are detailed in a formal model found in the Appendix.

The information observed by individuals in our studies is assumed to be non-ambiguous and from equally credible sources, thus we do not consider the coding biases related to these forms of misperception. Instead, we expect another bias in the coding step to be a major driver for the bracketing effect. More specifically, we propose that the evaluation made at the coding step is based on a psychological likelihood judgment that resembles representativeness (Tversky and Kahneman 1971). The observer evaluates the new partition of information by considering whether this partition is representative of the underlying process that may be generating it, relative to other possible underlying processes. We expect that as individuals generate this “relative representativeness” judgment, both the proportion of positive to negative signals and the size of the sample will be taken into account. In other words, observers will consider 1) which of the available outcome alternatives have base rate proportions most similar to the proportions in the sample, and 2) whether they think the sample is large enough to be meaningful. Smaller samples (narrow brackets) will look less convincingly representative to the observer, resulting in a representativeness judgment that is less extreme than for a larger sample with the same proportions.

This idea that small samples can be less representative than larger samples varies somewhat from the exact definition of representativeness provided by Tversky and Kahneman. According to Tversky and Kahneman’s (1971) original study of representativeness, small samples are expected to both appear random in their pattern, and locally resemble the underlying distribution subjects were participating in.
of the population from which they are drawn. While we expect the same result, our proposed representativeness judgment includes an increased sensitivity to sample size for very small samples, of less than perhaps four observations. Moreover, we expect that this representativeness judgment will not be a linear function of sample size, or even necessarily strictly monotonically increasing. Instead, the function may be S-shaped, such that there is little difference in the representativeness judgments of very small samples (e.g., under 3 signals) and also little difference in the representativeness judgments of larger samples (e.g., greater than 5 or 6 signals), but some increasing function in between these extremes. What this definition of representativeness suggests is that people do have some grasp of the law of small numbers (i.e., a single observation is not enough evidence for a hypothesis), but that their definition of a small sample is much smaller than a statistician’s, just as Kahneman and Tversky originally demonstrated.

We can see more clearly how representativeness might be influencing the beliefs being formed under the different bracketing levels by looking back to the example provided earlier. When narrowly bracketed into partitions of two observations each, no single partition (e.g., ‘11’) provides enough information to be judged as highly representative of either urn. However, when the bracket size is broadened to 5 (e.g., ‘11101’), then these partitions are more representative of a particular urn. Since the larger partitions seem more representative of the underlying process, the broader bracket’s representativeness judgment will be more extreme. Note that the overall representativeness of two partitions when they are combined into one may be very different than each partition’s individual representativeness measure.

The second departure from Bayes’ rule in our model is in the integration step. Our model is consistent with both Edwards’ conservatism model and the averaging models that followed it,
with one notable exception. Although Edwards took the conservatism factor to be a constant, implying that all new observations were equally weighted during integration, we predict that there will be order effects in how different signals are integrated into the overall judgement. More specifically, we expect both a primacy and a recency effect, such that the first and last partitions of information will be more heavily weighted than the partitions that occur in the middle. Primacy effects in these types of tasks have been documented previously (see, e.g., Hogarth and Einhorn 1982). In a sense, this can be considered a form of anchoring; the initial observation serves as the baseline for later judgment, and thus receives the heavier weighting. The role of the recency effect is less clear and is not always observed in these tasks (Hogarth and Einhorn 1982 have shown it to hold for certain tasks). We expect a recency effect because it is salient to the observer in our studies which observations are the final ones. These last observations coincide with the opportunity to make a final adjustment to the overall judgment, and are thus overweighted. The overall effect is that the integration step appears to average the partition-level judgments, with heavier weight accorded to the first and last partition. The coding and integration process for this model as applied to information received in brackets is depicted in Figure 3.³

³ Aspects of this process are evident in Mullainathan’s (2002) model of the categorization process. In his model, an observer judges the likelihood of a certain generating process based on how representative the
and that the bracketing effect may reverse in certain circumstances. Information delivered in smaller brackets will be judged less extreme than the larger brackets in cases where the initial and final partitions are more representative of the hypothesis than the overall pattern.

It is critical also that this model assumes that the judgment prior to seeing any particular signal is a summary statistic for the history up to that time, just as the prior in the Bayesian model is a summary of prior signals. Put differently, we assume that individuals do not reconsider the previous signals in incorporating the current signal, just as Ariely and Zauberman’s subjects do not reconsider previous experiences in judging their satisfaction. Of course, this does not influence judgments if individuals are Bayesian, since Bayesian updating obeys the product rule. However, our proposed model does not.

The result of the representativeness-based coding process and the order-specific updating process is that the final judgment of the overall sequence will be biased in predictable ways. When judging the observations generated by an unknown process, small subsets of observations will seem less representative, and therefore less diagnostic, of the underlying process. As a result, narrow bracketing will produce less extreme judgments. In contrast, observations shown in large subsets will be more representative, and thus more diagnostic, of the underlying process. Thus, broad bracketing will produce more extreme judgments. These two predictions are formalized in our first hypothesis:

**H1:** When judging the underlying properties of an uncertain process under **narrow bracketing**, small partitions will seem less representative, thus leading to **reduced extremity** in judgment. Similarly, when judging the underlying properties of an uncertain process under **broad bracketing**, large partitions will seem more representative, thus leading to **more extremity** in judgment.

---

observations are of that process. Once the initial category has been judged in this way, later observations can strengthen this category belief, or push the observer toward believing in a different category.
Although we expect that the majority of sequences will demonstrate this hypothesized bracketing effect, we also expect that the bracketing effect will be moderated for some sequences due to recency and primacy during the integration step. Since the initial and final partitions in a sequence are weighted heavier than other partitions according to our hypothesized model, having very representative partitions in these two locations could strongly influence the overall judgment for the sequence and possibly even reverse the overall effect. Although we expect such sequences of signals to be comparatively rare, we do allow for them with our second hypothesis:

**H2:** For sequences in which the initial and final partitions are particularly representative when observed under narrow bracketing, primacy and recency effects will moderate the overall bracketing effect and may result in more extremity in judgment.

These hypotheses are tested in a series of four studies.

**Studies of Belief Bracketing**

**Study 1: Restaurant Reviews**

Study 1 was designed to demonstrate the bracketing effect and to understand how the coding and integration processes operate under various bracketing conditions. To do so, we use binary signals and a continuous judgment measure. Study 1 looks at the bracketing effect in a natural environment. Our subjects saw restaurant reviews, and judged which restaurants would appeal to a friend. Although its primary focus is on judgment, this study also provides an opportunity to ask subjects about predicted behavior, to see whether changes in judgment affect intentions. Finally, although our overall prediction is that narrow brackets will lead to less extreme judgments than broad brackets in the majority of cases (Hypothesis 1), we also use Study 1 to introduce some tests of the second hypothesis by incorporating sequences of signals that are expected to lead to more extreme judgments under narrow bracketing.
Method

Participants

Participants in this study were 49 university students recruited through fliers posted on campus. Participants completed the 15 minute computer-based study in exchange for a payment of $4.

Materials and Procedure

The study was conducted on computer. Participants first read two pages of instructions which described the task. They then continued through a series of screens that provided restaurant reviews by independent critics displayed in the three different bracketing conditions of 2, 6, or 12 reviews at a time. We used a 3 (bracket size) × 21 (patterns) between-subject design. The 21 patterns were hand-created based on the types of patterns that we expected would be most sensitive to bracketing effects, 16 of which were created to be consistent with the prediction of more extreme judgments for broader brackets (H1), and 5 of which were designed to be more extreme when judged in narrow brackets (H2). An additional 9 patterns were judged in the 6-review partition condition, with the goal of collecting initial judgments for all possible partitions contained within the other 21 patterns. Thus, each participant judged a total of 24 patterns: 7 in the 2-review partition condition, 10 in the 6-review condition, and 7 in the 12-review condition. The 21 main patterns were balanced across the 3 bracketing conditions. No subject saw exactly the same pattern twice; however, several of the patterns were binary inversions of other patterns

4 Although we did not have a precise rule for generating these constructed patterns, our hypothesis was that the representativeness of the initial and final partitions would be the strongest influence on overall beliefs. Thus, for the 16 patterns constructed to match H1, the outer partitions were often designed to be less representative (e.g., ‘01’ or ‘10’), while the inner partitions were more extreme (e.g., ‘11’). For the 5 patterns constructed to match H2, the outer partitions were more extreme, while the inner partitions were less extreme and/or counteracted the extremity of the outer ones (e.g., initial partition ‘11’ counteracted by inner partition ‘00’).
(e.g., “0110” was inverted to “1001”).

The dependent variable for this study was a judgment of how likely a friend was to enjoy each restaurant. Subjects were given instructions that explained that they would be evaluating 3 online restaurant review systems. The systems differed in how the information was displayed; they were told that the systems could display critics’ reviews 2 at a time, 6 at a time, or 12 at a time, based on system data constraints. Reviews were binary, displayed as a “thumbs up” or “thumbs down” for each critic. The actual task was to predict, after each partition, the probability that a certain friend would like or dislike the restaurant under review. This measure was collected on a sliding scale that ran from 0% to 100%. Additional instructions specified that the critics’ reviews were all independent and provided in no specific order, and that the friend agreed with each critic’s review 60% of the time (diagnosticity level). Bracketing condition order was randomized. For each condition, the participant made their judgment about their friend’s probability of liking the restaurant after each partition, for a total of 6 judgments per restaurant in the 2-review condition, 2 per restaurant in the 6-review condition, and only 1 per restaurant in the 12-review condition. After seeing all 12 reviews for a restaurant, and making their final judgment for that restaurant, participants were then asked how likely they would be to visit that specific restaurant themselves. They then received feedback on whether or not their friend actually liked the restaurant being reviewed. This feedback was provided to encourage accuracy and to make the task more challenging and fun for the participants. Feedback was created to be consistent with a purely Bayesian judgment based on the pattern of reviews. This same process was repeated for all 24 restaurants that each participant evaluated. Thus, each subject completed a total of 69 judgments (42 in the 2-review condition, 20 in the 6-review condition, and 7 in the
After completing all 3 bracketing conditions, participants were asked to rate the ease-of-use for each of the 3 systems, and then to indicate which system was their favorite.

**Results**

The single judgment point that all 3 bracketing conditions have in common is the one made after all 12 reviews have been seen. This provides us with 21 judgments per subject for analysis (one for each of the 21 patterns of interest, balanced across the three bracketing conditions). All probability judgments are converted to an extremity measure by taking the absolute value of the judgment minus 50, the middle of the scale. Across all 21 patterns, the mean extremity measure for judgments is 15.0 in the 2-review bracket condition, 20.5 in the 6-review brackets, and 19.0 in the 12-review brackets. A repeated-measures ANOVA of these results shows a significant effect of bracketing ($F(2,144)=6.54$, $p<0.002$). Results per pattern are shown in Table 1.

Focusing on the 16 patterns designed to fit the Hypothesis 1 prediction, the result is equally strong; the 2-review condition is again less extreme ($M=15.9$) than the 6 ($M=23.4$) and 12 ($M=22.3$) conditions ($F(2,144)=8.56$, $p<0.001$). The results for the 5 patterns designed to fit Hypothesis 2 are in the predicted direction; the 2-review condition ($M=12.0$) and 6-review condition ($M=11.1$) are more extreme than the 12-review condition ($M=8.8$). However, a repeated-measures ANOVA showed an insignificant effect of bracketing ($F(2,82)=0.95$, $p=0.39$).

To more directly test whether the 2 types of constructed patterns were different in their reaction to bracketing, a dummy variable was incorporated into the analysis to see whether there was an interaction between the pattern type and the bracketing effect. This interaction was significant.

---

5 Restaurant names were fictional, and pretested to be neutral so that all information about the restaurant would come only through the user reviews.


(F(2,226)=5.57, \( p=0.004 \)), suggesting that the two types of patterns are significantly different in how they respond to bracketing. The average extremity of judgment for the two sets of patterns is graphed in Figure 4.

A similar analysis of participants’ reported likelihood of visiting each restaurant shows a bracketing effect on predicted behavior; likelihood estimates were more extreme for broad brackets than for narrow brackets. To create the likelihood extremity measures, we took the difference of participants’ likelihood estimates from 3 (the midpoint of the scale). Likelihood measures for the 2-review condition were less extreme (\( M=0.89 \)) than for the 6 and 12 conditions (\( M=1.13 \) and \( M=1.10 \)) \( (F(2,144)=7.42, \ p<0.001) \), indicating that participants seeing reviews in broader brackets were more likely to either visit or avoid the restaurant, while those seeing the narrow bracket were more unsure. Thus, Study 1 supports Hypothesis 1 that judgments are more extreme in broad brackets and also provides evidence that this difference in judgments carries through to predicted behavior.

The next step in our analysis of the Study 1 data was to determine how each partition is weighted during the judgment integration step. This analysis allows us to test for the hypothesized primacy and recency effects, as well as whether a simple averaging model could explain the variation in judgment extremity between the bracketing conditions. We chose to use an averaging model in this analysis because, as cited earlier, previous work on conservatism has shown that such updating is best modeled by an averaging rule (Trautman and Shanteau 1976), even if the actual process used by subjects is closer to an adjustment strategy (Lopes 1985).

The first step in constructing the model was to isolate an initial judgment for each possible partition of 2 or 6 used in the 21 patterns. This was straightforward for partitions of size 2, since there are only 4 possibilities (‘00’, ‘01’, ‘10’, and ‘11’), each of which occurred in the first spot.
in multiple patterns. It was also straightforward for partitions of size 6 that occur in the first bracket in each pattern, since those judgments are by definition the initial judgment. However, for partitions of size 6 that occur in the second bracket but did not occur naturally in the first bracket, we ran an additional 9 patterns separately in only the 6-review condition. The initial judgments for each of the unique patterns were averaged across subjects to form a partition-level valuation. Then, for all of the 21 patterns, the actual final judgments were regressed on the ordered sequence of partition-level valuations, within bracketing condition. In other words, all judgments made after seeing all 12 reviews in the 2-review bracketing condition were regressed on the average values of the ordered partitions that make up each pattern. This was done for the judgment made after 6 reviews in the 2-review condition, and for the judgments made after all 12 reviews in the 2-review and 6-review conditions. We define the variable $\text{Part}_{b,i}$ to indicate a partition of size $b$ presented in the $i^{th}$ position in the sequence, with $i=f$ for the final overall judgment for the sequence. The resultant models are as follows:

$\text{Part}_{6,f} = .22\text{Part}_{2,1} + .22\text{Part}_{2,2} + .58\text{Part}_{2,3} \quad (R^2_{adj}=.96)$,

$\text{Part}_{12,f} = .13\text{Part}_{2,1} + .20\text{Part}_{2,2} + .12\text{Part}_{2,3} + .08\text{Part}_{2,4} + .35\text{Part}_{2,5} + .14\text{Part}_{2,6} \quad (R^2_{adj}=.92)$

and

$\text{Part}_{12,f} = .34\text{Part}_{6,1} + .68\text{Part}_{6,2} \quad (R^2_{adj}=.94)$.

We use these models to predict judgments for any given pattern, and thus can examine how well these predicted measures fit the actual judgments by the participants. For each pattern, the predicted value of the final judgment was calculated using the appropriate weighting model. These judgments were then converted to extremity measures in the same way that actual subject judgments were converted. A regression was run using the actual subjects’ extremity measures as the dependent variable and the predicted measures from the weighted models as the independent
variables. The model’s predicted measures are highly significant for extremity. Additional regression analyses indicated a significant effect of bracket size on (a) extremity of judgment, consistent with the previously reported analysis, and (b) the predicted measures from the weighted model. Finally, regressions indicated that the effect of bracket size became insignificant when both bracket size and the weighted model’s predicted measures are included in the analysis, while the predicted measures themselves continue to be significant. Results from this series of regressions are provided in Figure 5. These results suggest that bracketing does have a significant effect on judgment, but that the underlying process causing this bracketing effect is an order-specific, weighted combination of partition-specific judgments. The individual judgments for the partitions that make up each larger pattern can significantly predict the overall judgment for that pattern.

Discussion

The results of Study 1 support the predictions of our first hypothesis, showing that most judgments formed under broad brackets are more extreme than those formed under narrow brackets. In this study, the participants that observed restaurant reviews 2-at-a-time were less extreme in their predictions about whether a friend was likely to enjoy the restaurant, and were also less likely to express an intent to visit the restaurant themselves. These results support our first hypothesis. Study 1 also allowed us to test our second hypothesis, that some sequences of observations may result in more extreme judgment under narrow bracketing due to primacy and recency effects. Although H2 was not strictly supported, we did find a significant moderating effect: a comparison of these sequences to the other patterns in the study indicates that they did respond differently to bracketing, thus providing evidence that primacy and recency matter.

One potential criticism of Study 1 is that the process of collecting a judgment from
participants after every bracket could be causing the bracketing effect. In other words, the effect may be the result of the task methodology rather than a more general result about how people process information. To address this would require using a task similar to Study 1, but in which judgments are collected only after all information has been seen. This modification forms the basis for Study 2.

**Study 2: Restaurant Reviews without interim judgments**

Study 2 is identical to Study 1, except that it removes the explicit judgments made by subjects after each partition of information. Subjects now make only a single judgment after seeing all 12 reviews, regardless of bracketing condition. However, the information is still delivered according to the relevant bracket size. We consider two possibilities for what will happen to the bracketing effect: it will go away, or it will persist. If the effect is based on an anchoring and adjustment process, then eliminating the interim judgments and allowing the subject to make a judgment only after seeing all information should attenuate or eliminate the bracketing effect. Subjects will then base their judgment only on the information as a whole, and not on judgments made after each bracket. However, if subjects spontaneously make judgments after seeing each new partition of information, then it should not matter whether we explicitly ask for a judgment or not; the bracketing effect should persist. There is some evidence that individuals spontaneously make judgments of information even when they are not required to (Hastie and Park 1986, Russo, Meloy, and Medvec 1998). This is the result we expect.

**Method**

**Participants**

Participants in this study were 30 university students recruited through fliers posted on campus. Participants completed the 10 minute computer-based study in exchange for a payment
Materials and Procedure

All materials and procedure were identical to Study 1, except that instead of entering a prediction after each partition, participants progressed through all partitions in each pattern before providing their estimate. Thus, only one judgment per restaurant was made under all three bracketing conditions (bracket sizes 2, 6, and 12), for a total of 24 judgments per subject.

Results

Across all three bracketing conditions, the single judgment point per restaurant is the one made after all 12 reviews have been seen. As before, all probability judgments are converted to extremity measures by taking the absolute value of the judgment minus 50. Overall results are consistent with Study 1; mean judgments across all 21 patterns are less extreme in narrow brackets and are similar in value to the mean judgments from Study 1 ($M=15.2$, 17.0, and 17.8 for brackets of size 2, 6, and 12, respectively). This pattern of results also holds for the 16 patterns designed to be consistent with Hypothesis 1 ($M=17.8$, 21.1, and 21.3 for brackets of size 2, 6, and 12, respectively). An analysis of these 16 patterns shows that this bracketing effect is still marginally significant, such that judgments are less extreme in the narrow bracket size when compared to the combined results from the broad bracket sizes ($F(1,88)=2.89$, $p=0.09$). As in Study 1, the results for the 5 patterns designed to fit Hypothesis 2 are not significantly different ($M=6.5$, 4.4, and 6.5 for brackets of size 2, 6, and 12, respectively; $F(2,47)=0.61$, $p=0.55$). Results for all patterns are shown in Table 2 and are graphed in Figure 6. An analysis of participants’ reported likelihood of visiting each restaurant shows the same effect on predicted behavior; likelihood estimates were more extreme for broad brackets than for narrow brackets for the overall 21 patterns. Likelihood measures for the two-review condition were less extreme
than for the 6 and 12 conditions \((M=0.99\) and 1.08\) \((F(2,87)=2.84, p=0.06)\), indicating that participants seeing reviews in broader brackets were more likely to either visit or avoid the restaurant.

To test whether there is any substantial difference in the results of this study from the results of Study 1, the data from both studies was put into a single analysis with a dummy variable to differentiate the two. Results of this analysis indicate that there is no significant difference in the results between the two studies \((F(1,233)=2.13, p=0.15)\).

Discussion

The results of Study 2 continue to support our predictions that judgments will be more extreme in broad brackets, even when judgments are only made after all information has been seen instead of after each bracket. Judgments by participants in Study 2 are nearly identical to those in Study 1; both judgment and intended behavior measures are more extreme under broad brackets than under narrow brackets.

One possible concern with the results from Studies 1 and 2 is that the patterns used to capture the bracketing effect were constructed according to the elements of our model that we predicted would lead to the desired effects. While this was necessary to distinguish between Hypotheses 1 and 2 (by providing us a priori with patterns that we expected would react to bracketing in different ways), the question of which effects are more prevalent in “real” random sequences has not been addressed. A second concern is that Studies 1 and 2 exploit memory biases; previous partitions are not retained in view while new partitions are presented. We expect that reducing the reliance on memory will not change the overall effect, since subjects will still prefer to just update the summary statistic of their prior judgment rather than reconsider all the data to create a new judgment from scratch. This is a modification that we test in the next study.
So far, Studies 1 and 2 have documented belief bracketing effects, using the restaurant review context. While high in external validity, these tasks do not have a normative standard with which to compare their results, thus limiting our ability to determine whether bracketing has an effect on judgment accuracy in addition to judgment extremity. A task similar to the traditional “bookbag and poker chip” studies provides a more direct test of bracketing’s effect on accuracy, as well as a test of some of the specific features of our model.

Study 3: Balls and Urns

We have demonstrated the bracketing effect in different hypothetical contexts in Studies 1 and 2. In Study 3, we conduct a more direct test of the model proposed earlier by testing it in a more traditional context. We use a computer-based experiment which tests the effect of bracketing on estimates of which of two possible urns a series of balls are drawn from. Study 3 also addresses other concerns from earlier studies by using randomly generated patterns rather than constructed patterns and by eliminating any possible effects of memory. As in earlier studies, we expect judgments in the most broad bracketing condition to be more extreme than judgments in narrow bracketing conditions. This experiment also provides us with a normative standard of judgment against which to check subjects’ performance so that we can see whether bracketing also has an effect on accuracy of judgment.

Method

Participants

Participants in this study were 44 university students recruited through fliers placed around campus. Participants were paid based on their performance in the study; actual payment ranged from $5 to $12.
Materials and Procedure

The study was conducted on computer and took approximately 45 minutes. Participants were given several pages of instructions that described the environment and their task. The basic task was to observe a series of 10 balls drawn from one of two urns; one urn was predominantly red, and the other urn was predominantly blue. Participants were asked to judge the probability that the current series of ten balls was being drawn from the blue urn. Each trial consisted of 10 draws (with replacement), all of which would be made from the same urn. This is a 3 × 3 design crossing 3 levels of bracketing (1 draw, 2 draws, or 5 draws per bracket) and 3 levels of diagnosticity (60%, 75%, and 90% colored urns). In the 1-bracket condition, participants made 10 estimates per trial, while in the 2-bracket condition, they made 5 estimates per trial, and in the 5-bracket condition, they made only 2 estimates per trial. These bracket sizes were chosen based on the Study 1 results, in which the bracketing effect was not significantly different between bracket sizes 6 and 12. For this study, we chose to not use the larger brackets, and instead focused on smaller brackets of 1, 2, and 5. The diagnosticity levels indicate whether the proportion of blue balls in the blue urn was 60%, 75%, or 90% (with probabilities reversed for the red urn). After reading the instructions, participants completed 2 practice trials. All participants then completed 9 1-bracket trials, 9 2-bracket trials, and 9 5-bracket trials, for a total of 27 trials (and 153 judgments) per participant. Within each bracketing condition, 3 of the trials were 60% diagnosticity, 3 were 75%, and 3 were 90%. Thus, each participant completed 3 trials from each cell of the design. Although the 9 trials within each bracketing condition were kept together (to reduce confusion by participants), the order of the bracketing conditions was counterbalanced, as was the order of the diagnosticity conditions. A break (with an unrelated filler task) was provided midway through the trials to minimize fatigue on later trials.
The actual sequence of balls in each trial was randomly generated in advance. There were 27 of these patterns created (9 for each diagnosticity condition), and each pattern was run in each bracketing condition. However, an individual subject never saw the same pattern twice. Thus, final analysis of bracketing effects per trial is between-subject. The fact that the 27 sequences were all randomly generated is important. Our previous restaurant review studies exploited our proposed representative-based judgment process and used sequences that were primarily constructed to be non-representative in narrow brackets and thus more likely to result in the desired effect. The use of random sequences in this study is a more stringent test; if most patterns do not differ in representativeness between small and large partitions, then no overall effect of bracketing will be found. Nevertheless, in Study 3, we expect that narrow brackets will produce less extreme judgments for the majority of the random patterns, consistent with Hypothesis 1. We also expect to find a small subset of patterns in which primacy and recency effects cause the bracketing effect to reverse such that narrow brackets results in more extreme judgment, consistent with Hypothesis 2.

Unlike Study 1 and 2, the full pattern was left visible to the observer as they made their judgments. Thus, all 10 balls in the pattern were still visible on the screen when subjects made their final probability estimate, regardless of the bracketing condition. This change helps control for a memory-based explanation for our earlier findings.

As incentive for accuracy, payment was calculated based on the difference between participants’ estimates and Bayesian estimates. Participants did not receive feedback during the task, and were unaware of their performance until the experiment was completed.

Results

Since our prediction is that broader brackets will lead to more extreme estimates, we
converted the subject estimates into an extremity measure. This measure indicates how far the subject is from the prior probability of 0.50. In addition, it is useful to have some indication of whether the subject’s estimate is extreme in the same or opposite direction of the Bayesian estimate. In other words, if the Bayesian estimate is 0.7, we would like to be able to distinguish between one subject’s estimate of 0.60 (which is 0.10 away from the prior) and another subject’s estimate of 0.40 (also 0.10 away from the prior, but on the opposite side of the Bayesian estimate). Thus, we used the following conversion rule to create our extremity measure:

\[
x = \begin{cases} 
0.5 - p & \text{if } p_{\text{bayes}} \leq 0.5 \\
p - 0.5 & \text{if } p_{\text{bayes}} > 0.5 
\end{cases}
\]

where \( x \) is extremity, \( p \) is the subject’s estimate of the probability that this pattern is coming from the blue urn, and \( p_{\text{bayes}} \) is the Bayesian posterior probability that this pattern is coming from the blue urn. Since the probability estimate for the \( 10^{th} \) observation on each trial is the only observation common to all 3 bracketing conditions, we use the \( 10^{th} \) observation from each pattern as the primary basis of analysis for the bracketing effect, leaving us with 27 judgments per participant for analysis.

Results for judgment extremity, by both bracketing condition and diagnosticity level, are shown in Table 3. The mean extremity for the broad bracket condition (bracket size of 5) is 0.30, which is larger than the mean extremity of 0.26 for both of the narrow bracket conditions (bracket sizes of 1 or 2). Looking only at the different mean extremity measures per bracket size and using a repeated-measures ANOVA to adjust for possible non-independence within each subject’s responses, we find a significant difference between brackets of size 1 and 5 (\( F(1,254) = 2.90, p=0.089 \)) and between brackets of size 2 and 5 (\( F(1,254)=3.51, p=0.06 \)), but not between those of size 1 and 2 (\( F(1,254)=0.09, p=0.76 \)). Since brackets of size 1 and 2 are so similar, we
collapse these results; this collapsed measure for the narrow brackets is significantly different from the brackets of size 5 ($F(1,381)=3.97, p=0.047$). Diagnosticity (the ratio of red to blue balls per condition) is significant as a main effect in all analyses ($F(2,381)=45.91, p<0.001$), but the interaction of bracket size and diagnosticity is not significant ($F(2,381)=0.46, p=0.63$). Overall, the results are consistent with our predictions, in that the broader bracket sizes lead to more extreme estimates by the subjects (i.e., probability judgments farther from the 0.50 prior).

To examine these effects in more detail, we consider some of the specific patterns that occur in Study 3. We start with pattern number 14, the same pattern used earlier as an example in the theory section. Recall from the example that when the sequence is seen in 2-ball brackets, each subset seems relatively undiagnostic on its own, and the overall evaluation of the sequence is not very extreme. However, when the sequence is seen under broader brackets of 5-balls each, it becomes very clear that the sequence is biased toward 0’s (in this case, toward blue balls). These results hold true for the data collected in Study 3. When seen with a bracket size of 2, subjects have an average probability estimate of 0.32 that the pattern came from the blue urn (resulting in an extremity measure of 0.18). However, when the 5-ball bracket is used, the probability estimate rises to 0.15 (extremity of 0.35), a difference in extremity that is significant. Thus, for this particular pattern, judgment extremity goes in the predicted direction, with broader brackets leading people to more extreme judgments.

However, it is also interesting to examine a pattern that leads to the opposite result, consistent with our second hypothesis. Remember that all 27 patterns were randomly generated, so the existence of such patterns is simply chance. One of these patterns is pattern number 6, which looks like ‘1011010011’. In this pattern, the subsets created in the 2-ball bracketing condition include three that seem more extreme, since they are either completely blue (‘11’) or
completely red (‘00’). Since the ‘11’ subsets outnumber the ‘00’ subsets, and the pattern ends with a ‘11’ subset that is heavily weighted due to recency, the pattern seen in these narrow brackets seems more extreme, and receives an average probability estimate of 0.74 for the blue urn, thus yielding an extremity measure of 0.24. However, under the 5-ball bracket condition, each 5-ball pattern has 3 1’s and 2 0’s. These two subsets do not seem very diagnostic individually, so the overall judgment is less extreme (probability estimate of 0.60, extremity measure of 0.10). Thus, in this case, broader brackets are actually less extreme.

Although our predictions specifically addressed the extremity of the subjects’ estimates, we can also look at the accuracy of the estimates as compared to the normative Bayesian estimate. To measure accuracy, actual Bayesian estimates were subtracted from the subject estimates to obtain a deviation term per observation. Again we restrict analysis to the 10th observation on each trial, the only observation that has an estimate from all three bracketing conditions.

Analysis of this data indicates a bracketing effect on accuracy; average deviation from the Bayesian standard is smallest in the 5-bracket condition (overall results shown in Table 4). Thus, broad brackets allow subjects to more accurately identify which urn is generating the sequence of draws. A regression of the deviations on bracket size and diagnosticity level shows that the bracketing effect is significant for bracket sizes one versus five ($t=-2.67, p<0.01$) and two versus five ($t=1.90, p=0.058$), but not for one versus two ($t=-0.77, p=0.44$). An ANOVA of the deviation on all observations also shows that the bracketing effect is significant ($F(2,1156)=3.32, p=0.036$). The effect of diagnosticity is also moderately significant for accuracy ($F(2,1156)=2.33, p=0.098$). This effect of bracketing on accuracy is not unexpected. Given the typical conservatism finding and our finding that estimates are more extreme for broader brackets, estimates that are more extreme should be closer to Bayesian and thus more accurate.
We return to this question regarding the effects of bracketing on judgment accuracy in the general discussion section.

Discussion

The traditional “bookbag and poker chip” paradigm used in Study 3 allows us to further investigate the effect of bracketing on probability judgments and then to extend our investigation to also consider the effect bracketing has on accuracy. Subjects who observed balls being drawn from an urn in brackets of 1 or 2 were less likely to provide extreme probabilities than those seeing balls drawn in brackets of 5 for the majority of random patterns. These subjects were also less accurate in their judgments for the more narrow brackets for those patterns. We were also able to identify some patterns in which the bracketing effect is reversed due to primacy and recency effects, consistent with our second hypothesis. All of these results are achieved without exploiting memory biases, since subjects continued to see all previous ball draws in the pattern while making their updated judgment.

Although this study confirms our predictions for both Hypotheses 1 and 2, and demonstrates that the primacy and recency effects are having some effect on the bracketed judgments, it does not explore the role that the representativeness function plays in the coding step in the process. Understanding representativeness helps us break apart the primacy and recency effects, enabling us to tell how much these initial and final partitions are weighted in the judgment process. Since probability judgments, and not explicit representativeness measures, were collected from the subjects in Study 3, we conducted a follow-up study to gather more data on which partitions and patterns in Study 3 were more or less representative of the hypotheses under consideration.

Study 4: Finding the Representativeness Function

Study 4 is designed to test the role that bracket-level representativeness judgments play in
the coding and integration process. Although Study 3 showed that bracketing does have an effect on the extremity of probability estimates, we can draw limited conclusions about the underlying psychological processes used. Study 4 was designed to shed light on the results from Study 3, primarily by examining the predicted role of the representativeness function. With this goal in mind, we collected representativeness measures for each version of the patterns used in Study 3.

Method

Participants

Participants in this study were 118 university students recruited in the student union and through flyers on campus. Participants completed the paper and pencil study in exchange for a payment of $2.

Materials and Procedure

Similar to Study 3, the study had three diagnosticity conditions (60%, 75%, and 90%). To reflect the large number of patterns in each diagnosticity condition, and to keep subjects from evaluating the same pattern in different bracketing conditions, there were two versions for each diagnosticity condition, resulting in 6 separate versions of the survey.

Subjects began with instructions that explained the idea of representativeness and the meaning of the diagnosticity level. The instructions then asked them to imagine that each series of balls were drawn from one of two urns, with replacement. Graphical examples of both a black and gray urn were provided, in which the diagnosticity level (ratio of black balls to gray balls) was apparent. Participants then proceeded through a series of patterns that presented draws in 2-
ball, 5-ball, and 10-ball formats. Each of the 6 versions of the study included two 2-ball patterns, seven (or eight) 5-ball patterns, and five (or four) 10-ball patterns. Subjects indicated both a black urn representativeness measure and a gray urn representativeness measure for each pattern using 11-point scales (with 10 labeled as “highly representative” and 0 labeled as “highly unrepresentative”), resulting in 2 representativeness measures for each of 14 patterns per subject.

Results and Discussion

We first analyze Study 4’s findings regarding representativeness, and then apply these findings to the Study 3 results. Consistent with our expectations, the representativeness measures are more extreme for patterns of length 5 or 10 than for patterns of length 2 (see Table 5). Also, as expected, representativeness measures are more extreme for the higher diagnosticity conditions (e.g., 90% black urns). Although we could further investigate what features of a pattern make it more or less representative, our primary interest with this data is to understand how the representativeness measures from the smaller brackets relate to the measures for the same patterns in larger brackets. Understanding how these representativeness measures relate may help us understand how the integration step operates in the belief formation process under bracketing.

To determine this relationship, we conducted an analysis similar to the one used in Study 1. We first regressed the 10-ball pattern representativeness measures on the measures collected for the relevant sub-patterns. Regressing the 10-ball pattern on the 2-ball patterns involves 5 independent variables (the regression was specified to have no intercept), while the 5-ball pattern regressions use 2 independent variables. The results of these regressions will provide us with a weighting model for how the narrow bracket measures are combined into the broad bracket measure, similar to the models built from the partition-level judgments in Study 1.
The findings for the model using the 2-ball partitions indicate that the heaviest (and most significant) weight is put on the last and first partitions, with the last partition having a weight of 0.81. In contrast, the third and fourth 2-ball partitions are non-significant, and the third partition even carries a negative weight. These coefficients suggest a strong recency and moderate primacy effect for the patterns, consistent with our model. The model using the 5-ball partitions also finds the heaviest weighting to be on the final partition, with a significant weight of 0.67. The first partition is also significant, with a weight of 0.35. Both models are robust, with respective $R^2$'s of 0.96 and 0.98. The models are:

$$\text{Part}_{10,f} = .17\text{Part}_{2,1} + .11\text{Part}_{2,2} - .15\text{Part}_{2,3} + .08\text{Part}_{2,4} + .81\text{Part}_{2,5} \ (R^2_{\text{adj}}=.96),$$

and

$$\text{Part}_{10,f} = .35\text{Part}_{5,1} + .67\text{Part}_{5,2} \ (R^2_{\text{adj}}=.98).$$

To connect these results from Study 4 to the extremity results from Study 3, we examine how well the representativeness function measures predict extremity. We ran a regression using the Study 3 extremity measures as the dependent variable and the outputs of the optimal weighted models for representativeness as the independent variables. The representativeness measures are highly significant for predicting the extremity measures. Additional regression analyses indicated a significant effect of bracket size on (a) extremity, as found in the Study 3 results, and (b) representativeness measures from Study 4. In addition, the regression indicated that the effect of bracket size became insignificant when the representativeness measures were added to the model while the representativeness measures themselves continue to be significant. The specific results of this series of regressions are provided in Figure 7.

What these results suggest is that bracketing does have a significant effect on judgment, as

---

7 A separate model with equal weights on each partition was also tested, and found to be useful, although with
shown in Study 3, but that the underlying process driving this bracketing effect is a bracket-specific representativeness function. The integrated model of the representativeness measures that make up the larger pattern can significantly predict the overall judgment of that pattern. Since shorter patterns (more narrow brackets) are typically judged to have less extreme representativeness measures than longer patterns (more broad brackets), the resulting judgments for a pattern observed in broad brackets will also be more extreme than those for the same pattern observed in narrow brackets.

**General Discussion**

The results of Studies 1-4 collectively indicate that bracketing of outcomes affects observers’ beliefs in predictable ways. Studies 1 and 2 used a restaurant review setting to demonstrate the bracketing effect, and to show that an averaging model for integrating judgments was consistent with the results. In addition, Study 2 tested whether the bracketing effect only occurred when interim judgments were required; the results of this study suggest that such judgments are spontaneously made, such that the bracketing effect still occurs when no explicit judgments are requested. Studies 3 and 4 provided a stricter test of the proposed belief formation model by examining representativeness judgments and their role in explaining how bracketing affects probability judgments. Study 3 also demonstrated that the predicted effect appears for naturally occurring sequences, that it is not dependent on memory (since previous observations were always displayed onscreen), and that bracketing also has an effect on judgment accuracy.

The effect of bracketing may be found in a variety of circumstances encountered by consumers and marketing managers alike. For example, the method by which the consumer collects information about product benefits or service reviews (as in the restaurant example) lower R² than the model presented.
could influence the beliefs formed by the consumer regarding whether the product or service will meet his needs. Similarly, the way in which the marketing manager delivers information about new benefits (as through advertising or product release schedules) will also affect the judgments made by the consumers about the overall value of those new benefits. Marketing managers themselves also risk being affected by bracketing effects in unexpected ways; for example, a salesforce manager reviewing recent performance reports by his subordinates may form different beliefs about their success depending on how often he reviews those reports, or how many reports he reviews at one time. Finally, the effect of bracketing on more specific types of judgment has yet to be explored, although the findings here raise some interesting possibilities. Consider judgments of risk; can the partitioning of outcomes of uncertain events affect an observer’s overall sense of how risky an option is? This could apply to low consequence judgments, such as what the chance of rain is today, as well as to high consequence judgments, such as which stock to invest in or whether to build a home in an earthquake zone. Just as broad brackets leads to greater extremity in the judgments we have investigated, it seems possible that broad bracketing will also result in more extreme assessments of risk.

Although we have primarily focused on extremity of estimates under varying bracket sizes, it is also interesting to consider the influence of bracketing on accuracy of estimates. Recall that an analysis of the subject estimates made in Study 3, when compared to the normative Bayesian posterior, indicated that subjects were most accurate (i.e., closer to the Bayesian standard) when in the broad bracketing condition. One explanation for this result is that broad bracketing leads to an increased perception of underlying trends, resulting in increased accuracy for overall judgments. For example, when an observer is trying to learn the underlying process for a series of data (such as balls being drawn from an urn), broad brackets provide a better view of actual
patterns and thus improve learning. However, there may be different types of tasks for which narrow bracketing encourages better accuracy, such as when underlying probabilities are known and events are independent. For example, consider an individual who is trying to predict the next flip of a fair coin after seeing a series of previous flips. Observers who treat each coin flip as independent through narrow bracketing are more likely to predict equal chance of heads or tails than observers who see the flips as part of a broadly bracketed whole and fall prey to a law of small numbers bias such as gamblers’ fallacy. This suggests that the optimal level of bracketing could depend on the uncertainty of the underlying process: when the underlying process is unknown and must be judged, then broad bracketing (increasing the amount of observations) will allow observers to better detect patterns, and thus lead to better accuracy. Conversely, narrow bracketing will obscure patterns in the data, and will lead to reduced accuracy. In contrast, when underlying probabilities are known and events are independent (as with a fair coin), then narrow bracketing may provide for better accuracy in prediction since it causes observers to focus on each event as independent and encourages observers to place more weight on the applicable base rates. Broad bracketing in these tasks will cause observers to perceive patterns that do not exist, and thus lead to reduced prediction accuracy. This question of the effect of bracketing on judgmental accuracy under these different environments remains a rich field for future research.
Appendix

Proposed Structural Model

The normative benchmark for the types of judgments we consider is Bayes’ theorem. Bayes’ theorem provides a precise formulation for both coding of signals and updating of beliefs. We review the Bayesian standard first, and then outline our proposed model and discuss how it differs from the normative standard.

Bayesian Standard

Let $H$ and $\bar{H}$ be the two specific hypotheses under consideration, each with prior probability $p(H)$ and $p(\bar{H})$, where $p(H) + p(\bar{H}) = 1$. We denote each signal as $x_i$, $i = 1, \ldots, n$, and a sequence of signals $x_1, \ldots, x_t$ as $X_{1, t}$. Signals are independent, and identically distributed, with the conditional probability of a signal $x_i$ given hypotheses $H$ and $\bar{H}$ denoted by $p(x_i | H)$ and $p(x_i | \bar{H})$, respectively. Bayes’ theorem provides the normative standard for updating beliefs upon seeing $X_{1, n}$. The posterior probability, in terms of odds-likelihood, takes the form

$$
\frac{p(H | X_{1, n})}{p(\bar{H} | X_{1, n})} = \frac{p(H)}{p(\bar{H})} \frac{p(X_{1, n} | H)}{p(X_{1, n} | \bar{H})}.
$$

(1)

Since the signals are i.i.d., (1) can be written in product rule form:

$$
\frac{p(H | X_{1, n})}{p(\bar{H} | X_{1, n})} = \frac{p(H)}{p(\bar{H})} \prod_{i=1}^{n} \frac{p(x_i | H)}{p(x_i | \bar{H})}.
$$

(2)

For our purposes, it is useful to separate two steps: coding and integration. Suppose that our observer has already seen $X_{1, n}$ and now receives a new signal $x_{n+1}$. A Bayesian “codes” the new signal by the likelihood ratio, $p(x_{n+1} | H) / p(x_{n+1} | \bar{H})$, and “integrates” the signal by multiplying this ratio and the prior likelihood ratio, $p(H | X_{1, n}) / p(\bar{H} | X_{1, n})$:
This process of coding and integrating a new signal is diagrammed in Figure 2. As the new signal is first evaluated (coded) and then integrated, the resulting posterior probability becomes the prior for any new signals subsequently received. Thus, the new prior serves as a sufficient statistic for the history of what the observer has seen. It is not necessary to remember the exact content or order of the previous signals. It is easy to see that, with this process, the order and/or bracketing of the signals have no effect on the overall posterior probability of a hypothesis.

Quasi-Bayesian models

Edwards (1968) found that subjects systematically violated Bayesian updating in numerous versions of the “bookbag and poker chip” studies. A consistent finding was that subjects were too conservative in their updating process. To model these departures from Bayesian updating, Edwards and other researchers considered models that were Bayesian in structure, but departed from Bayesian updating in how the elements were weighted and combined. We call these models Quasi-Bayesian models. Edwards adapted the Bayesian updating function (2) to reflect conservatism by raising the likelihood ratio by a constant factor, $c$, called an accuracy ratio. If $c < 1$, signals are discounted such that the final judgment is less extreme than the Bayesian standard. Edwards’ model is thus:

$$\frac{p(H \mid X_{1..n})}{p(\bar{H} \mid X_{1..n})} = \frac{p(H) \prod_{i=1}^{n} \left[ \frac{p(x_i \mid H)}{p(x_i \mid \bar{H})} \right]^c}{p(\bar{H}) \prod_{i=1}^{n} \left[ \frac{p(x_i \mid \bar{H})}{p(x_i \mid H)} \right]^c},$$

where Bayesian updating is the special case where $c = 1$. This model assumes no order effect, i.e., the accuracy ratio is constant across all signals.
Our proposed model builds on both the normative standard and the extensions described by Edwards, but takes a further step by incorporating bracketing effects. Our model departs from the traditional Bayesian standard in two main areas, both of which contribute to the hypothesized bracketing effect. First, there is a violation in the coding step, in which subjects use a “representativeness” function that differs from Bayesian likelihood, and does not in general obey the product rule. Second, there is a violation in the integration step, in which the update is both conservative (as Edwards found) and also sensitive to order effects.

More specifically, for the coding step for signal $x_i$, we replace the likelihood judgment with the signal valuation function $R(x_i; H)$, which indicates the “representativeness” of signal $x_i$ to hypothesis $H$ rather than $\bar{H}$, or $R_H(x_i)$ for simplicity. In the case of multiple signals observed as a single pattern, the function becomes $R_H(X_{i,j})$. We suggest that this representativeness function $R_H(X_{i,j})$ is sensitive to sample size effects. Since larger partitions are more likely to look strongly representative (or strongly not representative) of the underlying process, the resultant representativeness judgments will be more extreme. Thus, the overall representativeness of two samples when they are combined may be very different than each sample’s individual representativeness. In particular, we predict that larger samples will appear more representative than smaller samples, which can be formally represented as

$$R_H(x_i) \cdot R_H(x_j) \leq R_H(X_{i,j}), \text{ if } R_H(X_{i,j}) > 1$$

and

$$R_H(x_i) \cdot R_H(x_j) \geq R_H(X_{i,j}), \text{ if } R_H(X_{i,j}) < 1$$

The second departure from Bayes rule in our model is in the integration step for signal $x_i$, and
involves replacing the accuracy ratio \( c \) with an order-specific accuracy ratio \( \alpha_i \). Although Edwards took this factor to be a constant, we predict that there will be order effects in how different signals are integrated into the overall judgement. More specifically, when signals are observed one at a time, we expect both a primacy and a recency effect, such that \( \alpha_1 \geq \alpha_i \) for \( i \neq n \) (primacy) and \( \alpha_n \geq \alpha_i \) for \( i \neq 1 \) (recency). When signals are observed in brackets of size \( b \), for a total of \( s = 1 \) to \( n/b \) brackets, then we have \( \alpha_1 \geq \alpha_s \) for \( s \neq n/b \) (primacy) and \( \alpha_{n/b} \geq \alpha_s \) for \( s \neq 1 \) (recency).

Thus, in comparison to the normative standard (2) and the conservatism model (4), our proposed model instead takes the form

\[
\frac{p(H \mid X_{1 \ldots n})}{p(H \mid X_{1 \ldots n})} = \frac{p(H)}{p(H)} \cdot \prod_{i=1}^{n} \left[ R_H(x_i) \right]^{\alpha_i}.
\]

This form is readily extended for bracket sizes of \( b \):

\[
\frac{p(H \mid X_{1 \ldots b}, \ldots, X_{(n-b+1) \ldots n})}{p(H \mid X_{1 \ldots b}, \ldots, X_{(n-b+1) \ldots n})} = \frac{p(H)}{p(H)} \cdot \prod_{i=1}^{n/b} \left[ R_H(x_{(ib-b+1) \ldots ib}) \right]^{\alpha_i}.
\]

The coding and integration process for this model, especially as applied to information received in brackets, is depicted in Figure 3.

---

8 Note that the representativeness function ranges from 0 to infinity, as does the likelihood ratio used for the coding step in the Bayesian model. When the representativeness measure is greater than 1, the information favors hypothesis \( H \), and when it is less than 1, it favors the alternate hypothesis.
It is critical also that this model assumes that the judgment prior to seeing any particular signal is a summary statistic for the history up to that time, just as the prior in the Bayesian model is a summary of prior signals. Formally,

\[
\frac{p(H \mid X_{1,b}, \ldots, X_{(n-b+1),n})}{p(\overline{H} \mid X_{1,b}, \ldots, X_{(n-b+1),n})} = \frac{p(H)}{p(\overline{H})} \prod_{s=1}^{n/b} \left[ R_H(X_{(sb-b+1),sb}) \right]^{a_s} \frac{1}{\prod_{s=1}^{n/b} R_H(X_{(n-b+1),n})}.
\]

Put differently, we assume that individuals do not reconsider the previous signals in incorporating the current signal. Of course, this does not influence judgments if individuals are Bayesian, since Bayesian updating obeys the product rule. However, our proposed model does not.
References


Hastie, Reid and Bernadette Park (1986), “The Relationship Between Memory and Judgment Depends on Whether the Judgment is Memory-based or On-line,” *Psychological Review*, 93, 258-268.


Figure 1: Bracketing example

Prior odds
\[ \Omega_i = \frac{p(H \mid X_{1:i})}{p(H \mid X_{1:i})} \]

Observe signal
\[ x_{i+1} \]

Coding
likelihood = \[ \frac{p(x_{i+1} \mid H)}{p(x_{i+1} \mid \overline{H})} \]

Integration
\[ \Omega_{i+1} = \Omega_i \cdot \frac{p(x_{i+1} \mid H)}{p(x_{i+1} \mid \overline{H})} \]

Posterior odds \( \Omega_{i+1} \) become prior for next signal

Figure 2: Bayesian model of coding and updating
Prior odds
\[
\Omega_i = \frac{p(H \mid X_{1,i}, \ldots, X_{i-b+1,i})}{p(\bar{H} \mid X_{1,i}, \ldots, X_{i-b+1,i})}
\]

Observe bracketed signals
\[
X_{i+1,..,i+b}
\]

Coding
represntativeness = \( R_H(X_{i+1,..,i+b}) \)

Integration
\[
\Omega_{i+1} = \Omega_i \cdot \left[ R_H(X_{i+1,..,i+b}) \right]^{1/b}
\]

Posterior odds \( \Omega_{i+1} \) become prior for next signal

Figure 3: Proposed bracketing model of coding and updating with brackets of size \( b \)

Figure 4: Study 1 bracketing effects, per type of pattern

H1 indicates the 16 patterns constructed to be consistent with the predictions of Hypothesis 1, while H2 indicates the 5 patterns constructed to be consistent with the predictions of Hypothesis 2. Extremity is defined as the distance from 0.50.
Figure 5: Study 1 analysis using weighted model

A regression of Study 1 extremity measures shows that bracket size has a significant effect. Additional regressions show that including a weighted model of interim judgments in the analysis mediates the bracketing effect. (Standard errors shown in parentheses.)

Figure 6: Study 2 bracketing effects, per type of pattern

H1 indicates the 16 patterns constructed to be consistent with the predictions of Hypothesis 1, while H2 indicates the 5 patterns constructed to be consistent with the predictions of Hypothesis 2. Extremity is defined as the distance from 0.50.
A regression of the Study 3 extremity measures shows that bracket size has a significant effect on extremity of probability judgments. However, weighted models based on the representativeness measures collected in Study 4 mediate the Study 3 results (on a per pattern basis). (Standard errors shown in parentheses.)

**Figure 7: Integration of results from Studies 3 and 4**

A regression of the Study 3 extremity measures shows that bracket size has a significant effect on extremity of probability judgments. However, weighted models based on the representativeness measures collected in Study 4 mediate the Study 3 results (on a per pattern basis). (Standard errors shown in parentheses.)
Patterns 1 through 16 were constructed to be consistent with the predictions of Hypothesis 1, while patterns 17 through 21 were constructed to be consistent with the predictions of Hypothesis 2. Extremity of judgment indicates the distance of the actual probability judgments from a prior of 0.5.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Extremity of judgment</th>
<th>Extremity of likelihood estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>10.5</td>
<td>26.6</td>
</tr>
<tr>
<td>2</td>
<td>12.3</td>
<td>17.6</td>
</tr>
<tr>
<td>3</td>
<td>12.5</td>
<td>26.8</td>
</tr>
<tr>
<td>4</td>
<td>17.2</td>
<td>29.3</td>
</tr>
<tr>
<td>5</td>
<td>10.7</td>
<td>21.2</td>
</tr>
<tr>
<td>6</td>
<td>20.7</td>
<td>15.0</td>
</tr>
<tr>
<td>7</td>
<td>5.8</td>
<td>21.6</td>
</tr>
<tr>
<td>8</td>
<td>23.1</td>
<td>32.8</td>
</tr>
<tr>
<td>9</td>
<td>12.1</td>
<td>18.5</td>
</tr>
<tr>
<td>10</td>
<td>17.8</td>
<td>21.8</td>
</tr>
<tr>
<td>11</td>
<td>21.9</td>
<td>23.0</td>
</tr>
<tr>
<td>12</td>
<td>23.8</td>
<td>28.5</td>
</tr>
<tr>
<td>13</td>
<td>14.8</td>
<td>26.2</td>
</tr>
<tr>
<td>14</td>
<td>13.3</td>
<td>10.7</td>
</tr>
<tr>
<td>15</td>
<td>15.6</td>
<td>21.8</td>
</tr>
<tr>
<td>16</td>
<td>23.4</td>
<td>33.3</td>
</tr>
<tr>
<td>17</td>
<td>12.7</td>
<td>8.9</td>
</tr>
<tr>
<td>18</td>
<td>14.4</td>
<td>7.1</td>
</tr>
<tr>
<td>19</td>
<td>9.8</td>
<td>15.1</td>
</tr>
<tr>
<td>20</td>
<td>9.5</td>
<td>19.0</td>
</tr>
<tr>
<td>21</td>
<td>13.7</td>
<td>5.8</td>
</tr>
<tr>
<td>Average</td>
<td>15.0</td>
<td>20.5</td>
</tr>
</tbody>
</table>

Table 1: Results from Study 1 (restaurant review study)
<table>
<thead>
<tr>
<th>Pattern</th>
<th>Extremity of judgment</th>
<th>Extremity of likelihood estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>22.9</td>
<td>22.4</td>
</tr>
<tr>
<td>2</td>
<td>16.0</td>
<td>20.3</td>
</tr>
<tr>
<td>3</td>
<td>19.8</td>
<td>25.6</td>
</tr>
<tr>
<td>4</td>
<td>25.1</td>
<td>26.3</td>
</tr>
<tr>
<td>5</td>
<td>15.8</td>
<td>22.0</td>
</tr>
<tr>
<td>6</td>
<td>14.3</td>
<td>12.7</td>
</tr>
<tr>
<td>7</td>
<td>12.7</td>
<td>16.3</td>
</tr>
<tr>
<td>8</td>
<td>19.6</td>
<td>32.4</td>
</tr>
<tr>
<td>9</td>
<td>17.1</td>
<td>21.4</td>
</tr>
<tr>
<td>10</td>
<td>16.8</td>
<td>21.0</td>
</tr>
<tr>
<td>11</td>
<td>13.0</td>
<td>17.8</td>
</tr>
<tr>
<td>12</td>
<td>25.7</td>
<td>24.4</td>
</tr>
<tr>
<td>13</td>
<td>16.6</td>
<td>17.8</td>
</tr>
<tr>
<td>14</td>
<td>8.8</td>
<td>6.8</td>
</tr>
<tr>
<td>15</td>
<td>19.0</td>
<td>22.4</td>
</tr>
<tr>
<td>16</td>
<td>22.5</td>
<td>28.6</td>
</tr>
<tr>
<td>17</td>
<td>9.0</td>
<td>7.7</td>
</tr>
<tr>
<td>18</td>
<td>4.0</td>
<td>1.8</td>
</tr>
<tr>
<td>19</td>
<td>9.1</td>
<td>5.8</td>
</tr>
<tr>
<td>20</td>
<td>5.9</td>
<td>5.3</td>
</tr>
<tr>
<td>21</td>
<td>4.7</td>
<td>1.5</td>
</tr>
<tr>
<td>Average</td>
<td>15.2</td>
<td>17.2</td>
</tr>
</tbody>
</table>

Table 2: Results from Study 2 (restaurant review study)

As in Study 1, patterns 1 through 16 were constructed to be consistent with the predictions of Hypothesis 1, while patterns 17 through 21 were constructed to be consistent with the predictions of Hypothesis 2. Extremity of judgment indicates the distance of the actual probability judgments from a prior of 0.5.
Table 3: Extremity results for Study 3 (balls & urns)

Extremity of judgment, as measured by calculating the difference between the subjects’ final judgments and the prior of 0.5. Diagnosticity indicates the proportion of red to blue balls in the urns from which balls are being drawn.

<table>
<thead>
<tr>
<th>Diagnosticity</th>
<th>Bracket size 1</th>
<th>Bracket size 2</th>
<th>Bracket size 5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>.14</td>
<td>.16</td>
<td>.19</td>
<td>.16</td>
</tr>
<tr>
<td>3</td>
<td>.26</td>
<td>.25</td>
<td>.31</td>
<td>.27</td>
</tr>
<tr>
<td>9</td>
<td>.40</td>
<td>.37</td>
<td>.40</td>
<td>.39</td>
</tr>
<tr>
<td>Total</td>
<td>.26</td>
<td>.26</td>
<td>.30</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Accuracy results for Study 3 (balls & urns)

Average deviation from Bayes (subject estimate minus Bayesian estimate) for the 10th observation on each trial, per bracket size and diagnosticity level.

<table>
<thead>
<tr>
<th>Diagnosticity</th>
<th>Bracket size 1</th>
<th>Bracket size 2</th>
<th>Bracket size 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>.070</td>
<td>.042</td>
<td>-.018</td>
</tr>
<tr>
<td>3</td>
<td>.074</td>
<td>.046</td>
<td>.003</td>
</tr>
<tr>
<td>9</td>
<td>-.010</td>
<td>.007</td>
<td>.008</td>
</tr>
</tbody>
</table>

Table 5: Representativeness results from Study 4

Representativeness measures were collected on a 0 to 10 scale, with 0 indicating “very unrepresentative of urn” and 10 indicating “very representative of urn”. Subjects provided more extreme measures of representativeness for the larger bracket sizes (patterns of size 10).