Journal of Experimental Psychology: General
2011, Vol. 140, No. 1, 1–13

0096-3445/11/$12.00 DOI: 10.1037/a0021946

Grouping Information for Judgments

Anuj K. Shah and Daniel M. Oppenheimer
Princeton University

Models of cue weighting in judgment have typically focused on how decision-makers weight cues individually. Here, the authors propose that people might recognize and weight *groups* of cues. They examine how judgments change when decision-makers focus on cues individually or as parts of groups. Several experiments demonstrate that people can spontaneously pack information into cue groups. Moreover, group-level weighting depends on how people assess similarity or how they think of categorical hierarchies.

Keywords: cue weighting, judgment, information aggregation

Every week, the Gallup organization conducts polls of Presidential job approval. Although these data are sometimes presented as overall approval/disapproval statistics, they are at other times broken down by demographic group, such as race. For example, approximately 1 year into office, polls showed that 42% of White respondents approved of President Obama, compared with 74% of non-White respondents (Gallup, 2010). Gallup also presents these data in a slightly different way, breaking down the non-White respondents into two groups: Black respondents (89% approval) and Hispanic respondents (73% approval), which account for most but not all of the non-White respondents. These distinct methods of grouping and aggregating data can yield different interpretations, judgments, and decisions.

The twofold grouping—White and non-White—might be interpreted as saying that one group roughly approves of the President, whereas the other group does not. However, the threefold grouping—White, Black, and Hispanic—leads to the impression that two of three groups generally approve of the President's work, a considerably more favorable outlook for him. The Gallup organization also groups data according to political identification, including six groups: liberal Democrats, moderate Democrats, conservative Democrats, Independents, liberal/moderate Republicans, and conservative Republicans. This breakdown means that half of the groups are aligned with Democrats, which might lead to different impressions than if data were presented in a threefold grouping (e.g., Democrats, Independents, and Republicans) where only one third of the groups are aligned with Democrats. Although these various forms of presentation are all accurate, they appear to suggest different truths.

Of course, it is not just polling agencies that need to decide how to group data. Individuals also face this challenge. When making quantitative judgments, people must decide which pieces of information, or cues, are relevant to the judgment at hand, as well as how to weight these cues. People can combine these cues in a number of ways. One common, relatively simple strategy would be to assign equal weights to each cue.

But consider a person evaluating the performance of the stock market based on the performance of stocks for Google, Microsoft, and Bank of America. The three stocks can be grouped into two sectors: software and banking. The equal weighting strategy described above does not give equal weight to each sector. Instead, the software sector will account for two thirds of the decision-maker's response, whereas banking will account for one third. Thus, the decision-maker might instead opt to first group the cues into their respective sectors and to then average the sector-level data. However, this strategy gives Google and Microsoft less weight individually than Bank of America.

Our stock analyst is therefore left with a question about which level of aggregation is appropriate (cf. Hogarth, 1989), stock-level or sector-level. More generally, individuals must often decide between drawing inferences based on cue-level and group-level data. At the cue level, all cues are regarded as independent pieces of information. At the group level, some cues are regarded as related pieces of information. It is clear that a cue-level weighting strategy can yield different estimates than a group-level weighting strategy. And yet the question of how decision-makers group and categorize cues has been touched on only briefly in the literature.

In this article, we explore how judgments differ when people use cue-level and group-level weighting strategies. We also examine why people might weight information at a particular level. Past research has focused primarily on the ways that people integrate information at the cue level. But if decision-makers consider information at the group level, then this can substantially shift the weights that are applied to individual cues. Models of cue weighting might therefore need to be amended to account for weighting changes that result from group-level processing.

Anuj K. Shah and Daniel M. Oppenheimer, Department of Psychology, Princeton University.

These experiments were completed as part of Anuj K. Shah's doctoral dissertation at Princeton University. Anuj K. Shah thanks the members of his dissertation committee (John Darley, Dan Osherson, Betsy Paluck, and Eldar Shafir) and gratefully acknowledges support from the Woodrow Wilson Society of Scholars at Princeton University. We also thank Devin Karbowicz, David Mackenzie, Ani Momjian, and Jeff Zemla for their assistance in collecting data.

Correspondence concerning this article should be addressed to Anuj K. Shah, Princeton University, Department of Psychology, Green Hall, Princeton, NJ 08544. E-mail: akshah@princeton.edu

A Review of Weighting Strategies

There are a number of ways that decision-makers can reduce the complexity of judgment tasks and chief among them is to simplify cue-weighting strategies (Shah & Oppenheimer, 2008). Yet researchers have only identified a small, homogeneous set of strategies for weighting information. On the most complex extreme are strategies which assume that decision-makers weight cues according to validity (Gigerenzer, Todd, & the ABC Research Group, 1999; Payne, Bettman, & Johnson, 1993). On the other hand, the simplest strategies assume that decision-makers weight cues equally or randomly (Dawes, 1979; Gigerenzer et al., 1999).

For several reasons, validity-based models appear to be incomplete descriptors of human behavior. People certainly *can* weight cues according to validity, particularly when given time to learn validities (Slovic & Lichtenstein, 1971; Summers, 1962). However, learning is often slow and depends on the type of feedback given (Peterson & Pitz, 1985; Todd & Hammond, 1965). Moreover, learning is susceptible to a number of errors, such as those that result from competition between cues and the order in which cues are presented (Gluck & Bower, 1988; Slovic & Lichtenstein, 1971). There is also considerable evidence that people do not always use objective cue validity (e.g., Tversky & Kahneman, 1974) or even subjective cue validity (Evans, Clibbens, Cattani, Harris, & Dennis, 2003; Permut, 1973) when weighting information. A more complete understanding of cue weighting therefore needs to include models which are not based on validity.

Along these lines, Dawes (1979) presented a simple model where cues are equally weighted. This work initially emphasized the power and accuracy of linear models for numerical prediction, rather than whether these strategies were descriptive models of human behavior. Yet researchers have demonstrated that equal weighting often approximates how people approach numerical prediction (Lichtenstein, Earle, & Slovic, 1975) and judgments in other domains, such as consumer behavior (Troutman & Shanteau, 1976) and impression formation (Anderson, 1965).

If validity-based models represent one extreme and improper linear models (e.g., equal or random weighting strategies; Gigerenzer et al., 1999) represent the other, then there has been scant research on models of cue weighting that fall in between. To address this gap in the literature, we propose that decision-makers sometimes build more nuanced versions of simple weighting strategies. That is, simple weighting strategies need not always be so simple.

Group-level cue weighting might be one general example of such a strategy. When decision-makers identify cues that are from the same category or group, they can combine these cues to form group-level evaluations. These group-level evaluations can then inform the target judgment. The normatively correct level of aggregation will often depend on the task environment. We expect that decision-makers will be sensitive to these normative factors but that decision-makers will also adjust the level of weighting in response to other factors, such as the salience of similarity dimensions or levels in a categorical hierarchy. We focus less on normative concerns and more on the latter factors to highlight how cue grouping is a strategy that people can adopt independent of environmental demand.

The cue-grouping strategy was partially suggested by Hogarth (1989) as a "levels of aggregation" problem. Hogarth presented a vignette where a decision-maker assesses the advancement of an adversarial military based on reports from four satellites, three refugees, and one spy. All satellites say that the enemy is approaching. One refugee says that the enemy is approaching and two say the enemy is not. And the spy says the enemy is not approaching. Hogarth notes that this can be taken to mean that two of the three *types* of sources indicate that the enemy is not approaching, and so there is no need for alarm. Or the uncategorized data can be taken to mean that five pieces of evidence say the enemy is approaching, whereas three pieces of evidence say the enemy is not, leaving cause for alarm. There are few if any empirical tests of this brief note. In related work, Louviere (1984) developed a method for modeling cue weights based on the *assumption* that people might integrate information at multiple levels. This work focused on designing paradigms that might simplify statistical modeling of cue weighting, but it did not examine whether decision-makers actually group cues or what the consequences of cue grouping might be. So although the literature offers the beginnings of a theory about how people group information, there remains much to be developed. It would be useful to start by considering past research on how decision-makers use other types of groups and partitions to make judgments and decisions.

Unpacking and Partitioning

Although there has been little study of group-level cue weighting, there is substantial literature on how partitions and groupings affect probability estimates, such as the work on "unpacking effects." Unpacking effects occur when people assign a higher probability to an event when the event is broken down into its components. For example, one can assess the probability that a playing card drawn from a deck will have a "red suit" or the probability that it will be a "heart or diamond." These are equivalent events, but naming the two suits unpacks the notion of a "red suit."

Unpacking effects have been empirically demonstrated in a number of contexts. For example, Fischhoff, Slovic, and Lichtenstein (1978) asked people to estimate the probability that various systems in a car might lead to the car not starting. Participants assigned probabilities to a list of possible causes which included the category "other causes." For some participants, the "other causes" category was unpacked into additional systems (e.g., fuel systems, the engine). When the "other causes" category was described in more detail, participants assigned higher probabilities to it.

To explain unpacking effects, Tversky and Koehler (1994) developed *support theory*, which states that the probability assigned to a focal hypothesis is calculated by considering the strength of evidence (i.e., support) for the focal hypothesis relative to the support for alternative hypotheses. It is important to note that people do not automatically represent a hypothesis in terms of all of its components. For example, people do not automatically think of "other causes" of car failure in terms of the various unnamed fuel systems. Therefore, support for a "packed" event or hypothesis will almost always be less than support for an unpacked event (where the components of a hypothesis are made explicit).

Support theory is quite successful in explaining unpacking effects, but it does not speak to how people group (or "repack") components of a category which have been explicitly laid out.

From the playing card example above, support theory does not make predictions about when people will spontaneously think of "hearts and diamonds" as "red suits." To better inform our predictions about cases such as these, we turn to research on how people partition and group components of categories (Fox, Bardolet, & Lieb, 2005; Fox & Clemen, 2005; Fox & Levav, 2004; Fox, Ratner, & Lieb, 2005; Fox & Rottenstreich, 2003; See, Fox, & Rottenstreich, 2006).

For example, Fox and Rottenstreich (2003) examined how decision-makers partition sets of events. They suggested that decision-makers construct a set of n possible events and then assign each event a probability that is biased toward $1/n$. They call this strategy anchoring on an ignorance prior, which is based on the principle of insufficient reason (for a discussion, see Hacking, 2001).

To test this phenomenon, Fox and Rottenstreich (2003) primed participants to think of different event spaces. For example, some participants were asked the probability that "the temperature on Sunday will be higher than every other day next week" (Fox & Rottenstreich, 2003, p. 196). A separate group of participants was asked the probability that "next week, the highest temperature of the week will occur on Sunday" (Fox & Rottenstreich, 2003, p. 196). Participants in the former group typically constructed a twofold event space, where either Sunday could have the highest temperature or it could not. Participants in the latter group constructed a sevenfold event space, where the highest temperature could fall on Sunday, Monday, Tuesday, and so forth. Participants' subjective probability estimates were higher when given the twofold partition instead of the sevenfold partition, suggesting that they anchored on the ignorance prior established by the question. Such partitions have also been shown to play a role in conditional probability estimates (Fox & Levav, 2004) and frequency judgments for past observations (See et al., 2006).

Fox, Ratner, and Lieb (2005) have also shown how partitioning can affect choice and allocation. In one experiment, participants decided how much of a budget to allocate to international and local charities. All participants saw one international fund and four local funds. Participants in the nonhierarchical condition were simply asked how much money they would like to allocate to different funds, facilitating a fivefold partition. Participants in the hierarchical condition were first asked how much money they would like to donate to the superordinate funds (international and local), facilitating a twofold partition. These participants were then asked to specify how to distribute those funds among the subordinate groups. Participants in the nonhierarchical condition donated more in total to local funds than to the international fund, whereas participants in the hierarchical condition more evenly split their donations among the superordinate groups. It therefore appears that participants treated each fund equally unless attention was brought to the fact that there were two broad classes of funds with unequal numbers of local and international funds.

In choice contexts, participants often diversify their selections based on partition labels. For example, a wine list organized by region leads people to select one wine from each region. On the other hand, a wine list organized by varietal leads people to select one wine for each grape. Thus, arbitrary means of organization can shift how decision-makers choose from identical sets (Fox, Ratner, & Lieb, 2005).

Although the findings above do not speak to how people weight or aggregate information, they do inform an approach to studying cue grouping in several ways. First, if participants' probability estimates and allocations are biased toward an ignorance prior, then cue weights might similarly change based on whether people group information. Participants who group cues could be biased toward equally weighting cue groups, whereas participants who do not group cues could be biased toward equally weighting individual cues. Second, if hierarchical reasoning can affect how people allocate resources, then perhaps it can also affect how people integrate information. Finally, if arbitrary partitions affect consumer choice, then it is possible that arbitrary group labels can affect how decision-makers weight objectively identical information.

Grouping Information for Judgments

Before discussing how decision-makers group information, an important distinction needs to be drawn. Cues have two components: a cue type and a cue value (Shah & Oppenheimer, 2009). When buying a computer, you might first think to examine its storage capacity—the cue type. While considering this cue type, you must assess how many gigabytes the hard drive holds—the cue value. More generally, cue types are dimensions or attributes that have diagnostic power for a judgment. Each of these dimensions will have a value on some scale. We propose that decision-makers will primarily group information on the basis of the similarity of cue types. The dimensions along which similarity comparisons are made might depend on the question being asked or other contextual factors (Medin, Goldstone, & Gentner, 1993).

To more fully understand the implications of cue grouping, suppose that you are predicting a student's overall scholastic performance on the basis of his performance in four classes: calculus, geometry, French, and ceramics. If you did not attend to the relative similarity between calculus and geometry, you would be likely to use cue-level weighting. However, if you did notice the similarity, you would probably group calculus and geometry together as part of mathematical aptitude, whereas French and ceramics might stand alone as linguistic and artistic aptitude, respectively.

For the sake of simplicity, suppose that decision-makers are biased toward equally weighting individual cues and groups of cues. If the total amount of weight one can allocate to cues is 1, then under cue-level weighting one would assign weights of approximately one fourth to each of the student's classes. But under group-level weighting, one would first assign weights of one third to each of the three categories of aptitudes (mathematical, linguistic, and artistic). Then one would allocate weights to the cues within each group. Because linguistic and artistic aptitudes only have one cue each, these cues would receive the weights of one third assigned to the group. Because mathematical aptitude has two cues, one might evenly split the weight of one third assigned to the group. The two mathematics courses would therefore receive weights of one sixth each. Thus, relative to cue-level weighting, group-level weighting might dilute the weight that some cues receive while augmenting the weights that other cues receive.

Several points are worth noting. First, we present this theory of group-level weighting as an *as-if* model. That is, this theory is better for predicting what judgments people will produce rather

than the exact processes that people will use to produce these judgments. Although there are a number of cognitive processes that might underlie group-level weighting, we primarily focus on identifying the conditions that lead people to use this strategy. Moreover, it is certainly possible that decision-makers use strategies other than equal weighting at the cue and group levels. In fact, it is unlikely that people divide weights perfectly equally. Instead, we wish to emphasize that people are merely biased toward weighting cues (and groups) equally, just as probability estimates for an event are biased toward $1/n$ for n possible events (e.g., Fox & Rottenstreich, 2003). As such, the notion of cue grouping is a guide for relative, rather than precise, predictions about numeric estimates.

Illustrative examples should help clarify the patterns of judgment we expect if participants use group-level weighting. In some experiments, participants were shown four cues. Two of the cues had high values, and two had low values. For instance, if participants were judging the overall scholastic performance of a student (i.e., the target judgment) based on his performance in four classes (i.e., the cue types), then a high cue value would be 90/100 and a low value would be 10/100.

There were four conditions of interest. In "no-grouping" conditions, cue types were dissimilar enough to make grouping unlikely and to encourage cue-level weighting strategies. Cue types in the no-grouping condition might be the student's performance in calculus, history, French, and ceramics. Suppose that the student earned 90s in calculus and history and 10s in French and ceramics. Cue-level weighting would assign weights of one fourth to each class, essentially averaging the scores to yield an overall performance estimate of 50.

In other conditions, there were two cue types that were similar to each other and two cue types that were dissimilar to all other cue types. Cue types might be the student's performance in calculus, geometry, French, and ceramics, where participants are likely to group calculus and geometry but to not group French and ceramics. In "grouped-high" conditions, the two similar cue types had high values, whereas the remaining cue types had low values. At the group level, only one group—math—would have a high value, whereas the remaining two groups—language and arts—would have low values. More concretely, the calculations might proceed as in Equation 1:

$$\frac{\frac{(90+90)}{2}+10+10}{3} \approx 37. \tag{1}$$

The high scores in math both receive weights of one sixth, whereas the low scores of language and arts both receive weights of one third. Consequently, estimates should decrease relative to the no-grouping condition. "Grouped-low" conditions were analogous to grouped-high conditions, except that the two similar cue types had low values and the remaining cue types had high values. Calculations might proceed as in Equation 2:

$$\frac{\frac{(10+10)}{2}+90+90}{3} \approx 63. \tag{2}$$

Estimates should therefore be higher than in the no-grouping and grouped-high conditions. In "grouped-mix" conditions, the similar cue types had conflicting values (i.e., one high, one low). More concretely, the student might have performed well in calculus but poorly in geometry, while also doing well in French and poorly in ceramics. The same principles behind Equations 1 and 2 would yield estimates of 50.

In some experiments, participants saw only three cues. There were either two cues with high values (and one cue with a low value) or two cues with low values (and one cue with a high value). Suppose that a student did well in calculus and French but did poorly in ceramics. Grouping would not occur and so all scores would be equally weighted, yielding overall performance estimates of approximately 63 or 37, depending on if the student performed well or poorly in two of the classes.

Unlike the four-cue experiments, there were no grouped-mix conditions in these experiments. When grouping was possible, similar cues always had similar values. The dissimilar cue had a disparate cue value. Suppose that the student did well in calculus and geometry but did poorly in ceramics. Now, cue grouping would occur as in Equation 3:

$$\frac{\frac{(90+90)}{2}+10}{2} = 50. \tag{3}$$

Thus, if participants did not group these cues, then their judgments should be strongly in line with the majority of information. If participants did group these cues, then one group would have a high value and one group would have a low value, regardless of whether most cue values were high or low. In such cases, judgments should be closer to the midpoint of the target scale.

In all experiments, participants were told that the information was selected at random and was not necessarily exhaustive of all the information that might be pertinent to a judgment. In the experiments presented below, we explore various factors that moderate whether and how people group information, such as experimenter-provided group labels, attention to different dimensions of similarity, and consideration of different categorization levels.

Experiment 1a: Spontaneous Grouping

The aim of this first experiment is to establish the basic phenomenon of cue grouping and to show how people might spontaneously recognize groups of information. Moreover, this experiment will demonstrate how group-level and cue-level weighting yield different judgments. Participants were shown hypothetical data about what percentage of different animals carried a certain bacterium. In some conditions, cue types were chosen to facilitate group-level weighting. For example, grouping might emerge if data were shown about two birds, one fish, and one amphibian (e.g., crows, ravens, bass, and frogs). In no-grouping conditions, cue types were chosen to make groups less apparent, suggesting four categories instead of three (e.g., presenting information about crows, bass, frogs, and rabbits). We predicted that participants' judgments would reflect cue grouping when cue types highlighted group membership.

Method

Participants. Eighty-eight participants (58 women, 30 men; M age = 33.84 years) were recruited from Amazon.com's Me-

chanical Turk (MTurk) service. Using the service's built-in screening criteria, only users who resided in the United States were permitted to participate. Past demographics surveys have found that 65% to 70% of MTurk participants from the United States are female. Moreover, participants are typically younger and more educated than the general population (for more detailed information on MTurk demographics, see Paolacci, Chandler, & Ipeirotis, 2010).

Participants were prevented from participating more than once by using the service's screening mechanisms along with automatic checks for repeat visitors within the survey itself. Participants were recruited only for this experiment and therefore answered only the judgment question described below and demographics questions (age and gender). This is true for all experiments presented in this article.

Design, stimuli, and procedure. Participants were asked to estimate the prevalence of a hypothetical bacterial strain ("Comobacter") among animals. Participants were provided with prevalence data such that the bacterium was common among two animals and rare among two animals. The cue types were animals drawn from the following categories: birds (crows, ravens), fish (bass, trout), amphibians (frogs, toads), and mammals (rabbits, hares). There were two grouped conditions, constructed by selecting two animals from the same category and two animals from distinct categories. In the grouped-high condition, the animals from the same category both exhibited a high prevalence of the bacterial strain (89% and 87%), whereas the remaining two animals exhibited a low prevalence (12% and 15%). In the grouped-low condition, the animals from the same category both exhibited a low prevalence, whereas the remaining two animals exhibited a high prevalence. In the no-grouping condition, one animal was selected from each of the four categories, and two animals exhibited a low prevalence, whereas two animals exhibited a high prevalence of the bacterium. Participants gave prevalence estimates in response to the question, "Based on the above information, what percentage of all animals in the United States carry Comobacter?"

Predictions for the no-grouping condition were based on a simple average, and predictions for the grouped-high and grouped-low conditions were based on Equations 1 and 2, respectively. From these calculations, we would expect judgments to equal 51% for the no-grouping condition, 38% for the grouped-high condition, and 63% for the grouped-low condition. Participants might not produce these exact estimates, but we expected their judgments to be lowest in the grouped-high condition, highest in the grouped-low condition, and in the middle for the no-grouping condition.

Selection and ordering of animals were randomized between participants, subject to the constraints described above. Cue types (i.e., the animal names) were all presented in a single row and subjects clicked on the names to view the bacterium data. Participants were asked to read through all of the information and then to estimate the percentage of animals in the United States that might carry the bacterium (see Figure 1a for an example of how

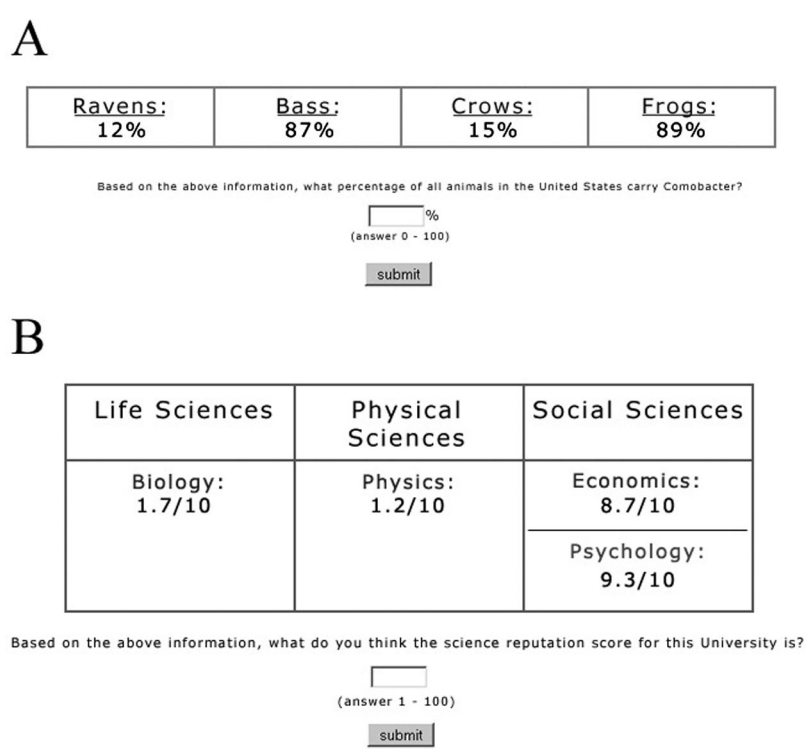

Figure 1. Examples of how stimuli were presented. A. In most experiments, participants were shown a table of cues which they needed to click to reveal the cue values. A grouped-low condition from Experiment 1a is shown. B. In Experiment 2, all information was organized by group labels, and all cue values were shown at once. A grouped-high condition is shown.

the stimuli were presented). Participants could not submit an answer until they had viewed the data for all animals.

Results and Discussion

A one-way analysis of variance (ANOVA) revealed a significant effect of condition on estimates of bacterium prevalence, $F(2, 85) = 3.26$, $p < .05$, $\eta^2 = .07$. Bonferroni-corrected post hoc analyses showed that participants in the grouped-high condition estimated the bacterium prevalence to be significantly lower ($M = 42.4$, $SD = 20.1$) than did participants in the grouped-low condition ($M = 55.3$, $SD = 24.0$), $p < .05$. Estimates in the no-grouping condition ($M = 50.8$, $SD = 13.9$) did not differ significantly from either of the grouping conditions. These data are consistent with the idea that participants used a cue-level equal weighting strategy when they based judgments on information about four dissimilar animals. However, when grouping was possible, participants' judgments were in line with the predictions based on group-level weighting.

But what makes cue types similar enough to merit grouping? One possibility is that there is an absolute standard for similarity between cues that leads to grouping. A more plausible explanation is that groups emerge on the basis of relative similarity and dissimilarity. Cue groups might become apparent when two (or more) cue types are similar to each other *and* dissimilar to other cue types. This notion is similar to what Tversky (1977) called the diagnosticity hypothesis for similarity comparisons. Experiment 1b was designed to address whether spontaneous cue grouping depends on a balance of similarity within cue groups and dissimilarity between cue groups.

Experiment 1b: Cue Grouping with Heterogeneous Information

In this experiment, participants estimated the nutritional information for the typical candy sold in the United States on the basis of nutritional information for three candies, which were selected from two groups. Consider two scenarios. In the first case, a participant sees information about Milky Way, Three Musketeers, and Snickers (all chocolate bars). In the second case, a participant sees information about Milky Way, Three Musketeers, and Twizzlers (two chocolate bars and a fruit-flavored candy). In both cases, the participant has the opportunity to group information about Milky Way and Three Musketeers. However, this grouping seems considerably more likely in the second scenario. That is, a heterogeneous set of cues makes the first two cues appear more similar, thereby increasing the likelihood of cue grouping.

Method

Participants. Forty-nine participants (36 women, 13 men; *M* age = 31.1 years) were recruited from MTurk.

Design, stimuli, and procedure. Whereas the previous experiment presented participants with four cues, this experiment presented participants with three cues. Participants were asked to estimate the sugar content of the typical American candy on the basis of information about the sugar content of three candies. Candies were selected from two groups: chocolates (Milky Way, Snickers, Three Musketeers) and fruity candies (Skittles, Starburst, Twizzlers). Cue sets could either be homogeneous (e.g., all chocolates) or heterogeneous (e.g., two chocolates and one fruity candy). And the majority of cues could either have high values (13 or 14 grams of sugar per serving) or low values (2 or 3 grams per serving). Thus, four conditions resulted from crossing these two between-subjects factors.

When cue sets were heterogeneous, the two similar candies were always given similar cue values, and the dissimilar candy was always given the disparate cue value. For example, a heterogeneous cue set with primarily high cue values might include the following sugar contents: Milky Way (13 g), Snickers (14 g), Skittles (3 g). Selection and presentation order of the candies were both randomized within these constraints. Participants gave their estimates in response to the question, "Based on the above information, how much sugar do you think the typical American candy contains in one serving?" Participants were told that their judgments should fall in the range of 1 to 15 g per serving.

We anticipated that participants would equally weight information in the homogeneous condition, because cue groups are less clear. Thus, we expected participants to produce judgments approximately equal to 10 g when the majority of cue values were high and 6 g when the majority of cue values were low. In the heterogeneous condition, however, we anticipated that participants would group cues. Following Equation 3, participants might produce estimates of approximately 8 g, regardless of whether most cue values were high or low. Once again, these calculations inform relative, rather than precise, predictions about participants' judgments. We expected that estimates in the homogeneous condition would be more disparate than estimates in the heterogeneous conditions.

Results

The response from one participant was excluded from the analyses as an outlier based on inspection of a boxplot, where whiskers were drawn using the common standard of 1.5 times the interquartile range. A 2 (Cue Set: homogeneous vs. heterogeneous) × 2 (Majority of Values: low vs. high) ANOVA revealed a significant main effect of the majority of values, $F(1, 44) = 131.75$, $p < .001$, $\eta_p^2 = .75$. It was not surprising that participants judged the typical candy to contain more sugar when most cue values were high rather than low. More critically, this main effect was qualified by a significant interaction between the two factors, $F(1, 44) = 5.56$, $p < .03$, $\eta_p^2 = .11$. As shown in Table 1, estimates in the homogeneous condition were more disparate than estimates in the heterogeneous condition.

Table 1
Experiment 1b: Average Judgments of the Sugar Content in the Typical American Candy, Shown as a Function of Cue Sets and the Majority of Cue Values

Cue set	Majority low	Majority high	Difference
Homogeneous	7.4 (2.1)	14.0 (1.0)	6.6
Heterogeneous	8.6 (1.5)	13.0 (1.7)	4.4

Note. Standard deviations are in parentheses.

Discussion

Clearly, participants paid attention to what the majority of cue values were. However, as expected from the cue-grouping hypothesis, this effect was significantly diminished under heterogeneous cue sets. It seems that groups which were not apparent under homogeneous cue sets became evident when a dissimilar cue was substituted into the set of information.

The results from these initial experiments suggest that participants gave less weight to similar cues than to dissimilar cues. This is consistent with the grouping hypothesis. But information was not held constant across conditions, so it is possible that the observed differences in judgments resulted from the change in informational content instead of from differences in *how* people weighted information. Related to this point, participants in some conditions were given a greater diversity of information (i.e., animals from four categories instead of three and candies from two categories instead of one). In the domain of induction, confidence in a conclusion often changes as a function of the diversity of the premises used to arrive at that conclusion (Osherson, Smith, Wilkie, López, & Shafir, 1990). The diversity of data points might similarly influence how people compute quantitative judgments, and this might not necessarily be related to cue grouping.

Finally, these studies permitted participants to use prior beliefs when making these judgments. Indeed, these studies depended on participants using prior associations as the basis for cue grouping. However, we also provided participants with fabricated data upon which to base their judgments. It is difficult to predict how these prior beliefs might interact with the fabricated data when such data are not held constant across conditions. Moreover, when cue types differ across conditions, this increases the likelihood that prior beliefs will also differ across conditions. All of these concerns center on the fact that information varied across conditions. It is therefore necessary to demonstrate that group-level weighting can change interpretations of identical data sets and to examine which factors might lead people to adopt cue-level or group-level weighting.

Experiment 2: Using Group Labels

This experiment tests whether arbitrary group labels might shift how decision-makers weight otherwise identical pieces of information. Participants were asked to rate the strength of university science programs on the basis of the quality of individual departments. The cues provided were the strengths of four individual departments. Three group-level labels were provided, organizing the departments by science area. We expected that participants would use a weighting strategy based on the group-level labels (i.e., science areas), rather than based on the cue types (i.e., individual departments).

Method

Participants. Two hundred participants (104 women, 96 men; *M* age = 31.3 years) were recruited from MTurk.

Design, stimuli, and procedure. Participants were told that they would be rating the strength of a university's overall science program based on assessments of four individual departments: physics, biology, economics, and psychology. These departments were organized into three science areas: physical sciences, life sciences, and social sciences. Physics was always listed under physical sciences, biology was always listed under life sciences, and economics was always listed under social sciences. Psychology, however, could be listed under the life sciences or social sciences. If participants grouped the information according to science area, then listing psychology as a life science would dilute the weight placed on biology, whereas listing psychology as a social science would dilute the weight placed on economics.

Participants were told that they were only being shown a sample of departments from a university's science program and that this sample was not necessarily representative of the science program's composition. Thus, if there were two departments listed under the life sciences and only one department listed under the physical sciences, that did not mean there were twice as many life science departments as physical science departments at the university.

There were eight between-subjects conditions that resulted from crossing three binary factors. The first factor was whether psychology was categorized as a social or life science. The second factor concerned the quality of the psychology department. Participants observed either a strong or weak psychology rating. The third factor concerned the quality of the department paired with psychology, again strong or weak. Strong departments were given scores of 8.7/10 or 9.3/10. Weak departments were given scores of 1.2/10 or 1.7/10. Department ratings were presented in a table with science areas written as column headings (see Figure 1b). The order of groups was randomized between subjects. Participants were asked to read through the information and to then answer the question, "Based on the above information, what do you think the science reputation score for this University is?" Participants were told to answer on a scale of 1–100, where higher numbers indicated a stronger reputation.

Grouped-high conditions refer to when a strong psychology department was grouped with another strong department. Grouped-low conditions refer to when a weak psychology department was grouped with another weak department. There were two grouped-mix conditions, which refer to when a strong psychology department was grouped with a weak department or when a weak psychology department was grouped with a strong department.

From the calculations described in the introduction, we expected that estimates of overall science reputation would be lowest when two strong departments were grouped together and highest when two weak departments were grouped together. We expected estimates to fall near the midpoint of the scale when strong and weak departments were grouped together. Note that if participants used a cue-level weighting strategy, then estimates would always be near the midpoint of the scale, because two of the departments were always strong and two of the departments were always weak.

Results and Discussion

The patterns of results described below did not differ on the basis of whether psychology was categorized as a social science or life science. Therefore, no further attention is given to that factor. Because assumptions of normality were violated, analyses used nonparametric statistics.

Participants in grouped-high conditions gave lower overall science ratings (M = 34.4, SD = 21.3) than did participants in grouped-low conditions (M = 57.3, SD = 20.4; Mann-Whitney

test $z = 4.79$, $p < .001$). Moreover, participants who saw a strong psychology department grouped with a weak department (a grouped-mix condition) gave higher overall science ratings ($M = 47.3$, $SD = 25.8$) than did participants in grouped-high conditions (Mann-Whitney test $z = 3.03$, $p < .003$). And participants who saw a weak psychology department grouped with a strong department (a grouped-mix condition) gave lower overall science ratings ($M = 47.2$, $SD = 25.8$) than did participants in grouped-low conditions, but this effect was not significant (Mann-Whitney test $z = 1.51$, $p = .13$). These results are consistent with the hypothesis that group-level labels alter participants' interpretations of the same cue-level data.

Participants behaved as if diluting the weight placed on biology and economics when these departments were paired with psychology. Consider an example where the biology department was strong, the psychology department was weak, the economics department was weak, and the physics department was strong. When psychology and economics were grouped together, estimates of overall science quality were more in line with the strength of the biology department. But when psychology and biology were grouped together, estimates of overall science quality were lower and therefore more in line with the weakness of the economics department. Thus, grouped cues appeared to receive less weight than ungrouped cues.

Experiment 2 demonstrates how cue grouping can influence the interpretation of almost identical data sets. In this case, participants were sensitive to experimenter-provided group labels. However, the results from this experiment might indicate that participants guessed the experimental hypothesis or were assuming that these labels were given because they should be used. Furthermore, the group labels might have provided participants with important semantic information—it might be useful to know whether a university considers psychology a life science or social science. Experiments 3 and 4 therefore test how judgments change when people are primed in more subtle ways to attend to different groups of information.

Experiment 3: Priming Dimensions for Grouping

When weighing information, decision-makers must often think about the different aspects of each cue. For instance, a car buyer can regard the information about fuel efficiency as an economic aspect as well as an environmental aspect. In Experiment 3, different dimensions of a cue were primed to change which groupings were salient. Participants in this experiment were asked to estimate the prevalence of a pest across the entire United States based on data for different cities. Each city could be considered in terms of multiple geographic dimensions. For example, cities might be considered northern or southern. They also might be thought of as eastern or western.

Participants were given information about three cities which were evenly distributed across the United States on the east–west axis. However, two cities were northern and one was southern. The data were structured such that the northern cities had similar prevalence rates (both were high or both were low), whereas the southern city had a disparate prevalence rate (low when the northern cities were high and high when the northern cities were low). Participants were primed to think in terms of an east–west or a north–south axis. A north–south prime should facilitate cue grouping, but an east–west prime should not, because multiple cities were not clustered along this axis. Therefore participants given the east–west prime should make judgments more in line with the majority of cue values, and participants in the north–south prime should make judgments closer to the midpoint of cue values.

This experiment also had a secondary goal of addressing whether decision-makers form group-level evaluations which might in turn inform other judgments. If participants form group-level evaluations, then these evaluations should inform judgments about novel group members. The north–south prime might therefore lead participants to heavily weigh information about the south when evaluating a southeastern city, whereas the east–west prime might lead participants to heavily weigh information about the east for the same evaluation.

Method

Participants. Seventy-two United States resident (45 women, 27 men; M age = 33.1) were recruited from MTurk.

Design, stimuli, and procedure. Participants were asked to estimate the prevalence of straw mites in the United States based on hypothetical information about how prevalent these mites are in three cities: Boston, Minneapolis, and Los Angeles. These cities were selected to be evenly distributed across the United States on the east–west axis but to create clustering when considered in terms of the north–south axis. The information was structured such that Boston and Minneapolis always had similar prevalence rates, whereas Los Angeles had a disparate prevalence rate.

There were four between-subjects conditions that resulted from crossing two factors. The first factor was whether participants were primed to think of an east–west or north–south axis. Upon entering the survey, participants were asked to indicate where they lived in the United States. In the east–west prime, participants were asked whether they lived closer to the east coast or west coast. In the north–south prime, participants were asked whether it typically snowed in the winter where they lived. Not all northern cities in the United States receive snow, but a snow/no-snow grouping is identical to a north–south grouping for these cities.

The second factor was whether the majority of information (i.e., the prevalence rates for Boston and Minneapolis) indicated high prevalence (86% or 89%) or low prevalence (5% or 8%) of straw mites. Information about the cities was presented in a single row and in a random order, as in Experiment 1a. Participants were required to view information for all cities before estimating the overall prevalence of straw mites in the United States. Predictions were based on calculations similar to those for Experiment 1b. Once again, we expected prevalence estimates to be more extreme if participants used cue-level weighting than if they used group-level weighting. That is, we expected estimates to be more extreme in the east–west prime than in the north–south prime.

Participants were also asked to evaluate mite prevalence for a novel city whose group membership could be construed differently depending on the prime. After participants estimated the straw mite prevalence for the entire country, they were then asked on a new page to estimate the prevalence rate for Miami, a southeastern city. Participants were not given any new information and were told to make this judgment based on the information that they had seen on the previous page.

Results and Discussion

The responses from three participants were excluded from the analyses for giving estimates that exceeded the range of provided data. Including these outliers in the analyses does not change the general pattern of results but slightly weakens the interaction of interest. A 2 (Majority of Cue Values: high vs. low) × 2 (Prime: east–west vs. north–south) between-subjects ANOVA revealed a significant main effect of the majority of information, $F(1, 65) = 30.65$, $p < .001$, $\eta_p^2 = .32$. Participants estimated that straw mites were more prevalent across the United States when the majority of cities had high mite prevalence instead of low mite prevalence. Critically, this main effect was qualified by a significant interaction between the majority of information and prime conditions, $F(1, 65) = 4.02$, $p < .05$, $\eta_p^2 = .06$ (see Table 2). Consistent with predictions, participants produced less extreme responses in the north–south conditions (where grouping was expected) than in the east–west conditions (where grouping was not expected).

Similar analyses were conducted for estimates of mite prevalence in Miami. To make the logic clearer, we will adopt slightly different terms for the "majority of cue values" factor. For this analysis, take the "majority high" conditions to indicate that the east (based on information about Boston) had high mite prevalence and the south (based on information about Los Angeles) had low mite prevalence; vice versa for the "majority low" conditions.

Because participants did not have any prior evidence about mite prevalence in Miami, they could have assumed that its prevalence rate was similar to the national average which they had just estimated. However, we expected that participants would consider Miami in terms of a group (or geographic region) which had been made salient by the prime given at the beginning of the experiment. That is, participants primed with the east–west axis would be more likely to think of Miami as an eastern city, whereas participants primed with the north–south axis would be more likely to think of Miami as a southern city. Therefore, participants in the east–west prime would produce estimates for Miami that were similar to what they had learned about Boston. And participants in the north–south prime would produce estimates for Miami that were similar to what they had learned about Los Angeles.

A 2 (Regional Prevalence: east-low, south-high vs. east-high, south-low) × 2 (Prime: east-west vs. north-south) between-subjects ANOVA revealed only a significant interaction, $F(1, 65) = 6.16$, $p < .02$, $\eta_p^2 = .09$. Simple-effects analyses showed that participants in the east–west prime estimated Miami to have significantly higher mite prevalence when they saw that the east had high prevalence and the south had low prevalence ($M = 60.5$, $SD = 34.8$) than when they saw that the east had low prevalence and the south had high prevalence ($M = 31.2$, $SD = 31.8$), $F(1, 65) = 5.15$, $p < .03$, $\eta_p^2 = .07$. However, participants in the north–south prime estimated Miami to have lower mite prevalence when they saw that the east had high prevalence and the south had low prevalence ($M = 33.2$, $SD = 36.5$) than when they saw that the east had low prevalence and the south had high prevalence ($M = 47.2$, $SD = 34.4$), but this difference was not significant, $F(1, 65) = 1.48$, $p = .23$, $\eta_p^2 = .02$.

Table 2
Experiment 3: Average Estimates of Straw Mite Prevalence Across the United States, Shown as a Function of Regional Prime and the Majority of Cue Values

Prime	Majority low	Majority high	Difference
East–west	25.1 (15.5)	55.9 (16.0)	30.8
North–south	29.2 (18.4)	43.6 (17.3)	14.4

Note. Standard deviations are in parentheses.

In previous experiments, cue grouping could have been attributed to changes in the information that was presented to participants. However, this experiment provides the first evidence that participants might group identical sets of information differently depending on the context established by primes. That is, when cues can be grouped according to multiple dimensions, decision-makers' estimates might change substantially when one of these dimensions is more salient than the other.

The secondary analysis presented here also offers some support for the notion that people might form group-level evaluations when grouping cues. When assessing the mite prevalence of Miami, participants appeared to use data about regions that were previously primed and were relevant. When the east–west axis was primed, participants appeared to use information from the eastern region in assessing Miami. When the north–south axis was primed, participants appeared to use information from the southern region. This evidence is, of course, not definitive. But these data are consistent with the idea that people evaluate groups (once they have grouped cues) and that they use these group-level evaluations to inform other judgments.

Just as decision-makers might attend to different dimensions for grouping, they might also be primed to interpret information at the cue-level or at the group-level. Many cues can be construed at different levels in a hierarchical organization, which can make some cue groups more or less apparent. As an analogy, think about zooming in or out on a map. When zoomed in on a map of the United States, you might only see a few states and not think of any immediate groupings. But if you zoom out slightly, regions emerge; the Midatlantic, New England, the Great Lakes. If you zoom out greatly, however, these state groupings are once again lost as you now attend to geography at the national level; the United States, Canada, Mexico. When weighting information, attention to the lowest or highest rungs on a hierarchy might similarly change or obscure certain cue groupings. We examined this possibility in the next experiment.

Experiment 4: Priming Attention to Group-level

In Experiment 4, participants estimated the nutritional content of the typical cereal based on the nutritional content of four cereals drawn from three groups. Participants were primed to think of foods either at a superordinate level (e.g., fruit, meat) or at a subordinate level (e.g., apples, bacon). We expected that participants who were primed to think of foods at the superordinate (categorical) level would be less attuned to similarities and differences between exemplars of the category. That is, at the superordinate level, cue types would only be thought of as examples of cereals, broadly construed. Put another way, all cereals would be folded into the same group or category. However, we expected that participants who were primed to think of foods at the subordinate (exemplar) level would be more accustomed to noticing similarities and distinguishing between specific types of products. There-

fore, at the subordinate level, chocolaty cereals and fruity cereals might not be folded into the same overarching group. Instead, these cereals might be thought of as belonging to separate *subgroups*. With the stimuli used in this study, we therefore expected cue grouping to occur after a subordinate prime but not after a superordinate prime.

It is important to note that there are probably cases where subordinate orientation would obscure cue groups, whereas superordinate orientation would make such groups more apparent. This experiment does not test the possibility that attending to subordinate features will *always* lead to cue grouping but, instead, tests the possibility that attention to different levels of a hierarchy can change how (or if) people group cues.

Method

Participants. Forty-two participants (29 women, 13 men; *M* age = 36.8 years) were recruited from MTurk.

Design, stimuli, and procedure. Participants were asked to estimate the carbohydrate content of the typical cereal sold in the United States on the basis of the carbohydrate content of four cereals. These cereals were drawn from three groups: fruity (Froot Loops, Trix), chocolaty (Cocoa Puffs, Count Chocula), and plain (Corn Flakes, Special K). The cue types were selected such that two were drawn from the same group and one was selected from each of the two remaining groups. Within these constraints, selection and presentation order of these cues were randomized between participants.

There were four between-subjects conditions that resulted from crossing two factors. The first factor was whether participants were primed to construe information at a superordinate or subordinate level (e.g., Fujita, Trope, Liberman, & Levin-Sagi, 2006). Participants in the superordinate prime were told, "Please name 10 types of foods you might find at a grocery store. Do not name specific products or brands, but instead name general types of food (e.g., 'soda' instead of 'Pepsi')." Participants in the subordinate condition were told, "Please name 10 specific products you might find at a grocery store. Do not name general types of foods, but instead respond with specific products or brands (e.g., 'Pepsi' instead of 'soda')."

The second factor was whether the cue values were distributed in a grouped-high or grouped-low condition. High cue values showed that a cereal contained 41% or 43% of the daily recommended carbohydrate intake per serving. Low cue values showed that a cereal contained 2% or 3% of the daily recommended intake. (In reality, these cereals contain approximately 8% per serving of the daily recommended carbohydrate intake; for example, see Nutrition facts, 2010.) Once again, participants were required to view all of the information before making their target judgment.

We did not expect cue grouping to occur under the superordinate prime. Because participants were primed to think of products at the categorical level, all cereals would probably be treated as members of the same group. The predictions for the superordinate prime therefore stem from calculations for no-grouping conditions. However, we did expect cue grouping to occur under the subordinate prime. Because participants were primed to think of specific products, they would likely notice similarities and dissimilarities between cereals. Instead of considering all cereals as part of the same group, participants were expected to consider multiple groups of cereals. The predictions for the subordinate prime therefore stem from the calculations for the grouped-high and grouped-low conditions, as described in Equations 1 and 2. We expected stronger differences in judgments across the grouped-high and grouped-low conditions for the subordinate prime than for the superordinate prime.

Results and Discussion

The response from one participant was excluded from the analyses as an outlier on the basis of inspection of a boxplot where whiskers were drawn to 1.5 times the interquartile range. A 2 (Grouping: grouped-low vs. grouped-high) × 2 (Prime: superordinate × subordinate) between-subjects ANOVA showed no significant main effects of either factor. However, there was a significant interaction between grouping and prime, $F(1, 37) = 4.55$, $p < .05$, $\eta_p^2 = .11$.

Simple-effects analyses revealed that participants who received the subordinate prime gave marginally higher target ratings in the grouped-low condition ($M = 27.4$, $SD = 5.9$) than in the grouped-high condition ($M = 21.2$, $SD = 7.2$), $F(1, 37) = 3.16$, $p < .09$, $\eta_p^2 = .09$. Yet in the superordinate prime, target ratings in the grouped-low condition ($M = 23.9$, $SD = 7.9$) were lower than the ratings in the grouped-high condition ($M = 27.8$, $SD = 8.4$), but this difference was not significant, $F(1, 37) = 1.48$, $p = .23$, $\eta_p^2 = .04$ (see Figure 2). These results show the usual grouping effect in the subordinate prime condition but not in the superordinate prime condition.

Although these data do not speak to whether superordinate or subordinate primes might generally lead to cue grouping, they do highlight how certain cue groups can become obscured as the level of construal changes. These results offer further evidence that decision-makers can interpret identical sets of information in substantially different ways depending on a primed mindset. Specifically, these results suggest that the extent to which people group cues might depend on whether they attend to hierarchical levels that accentuate group boundaries or obscure them by folding all cues into the same category.

General Discussion

In this article, we have suggested that previous models of cue weighting might be limited by considering only one level of

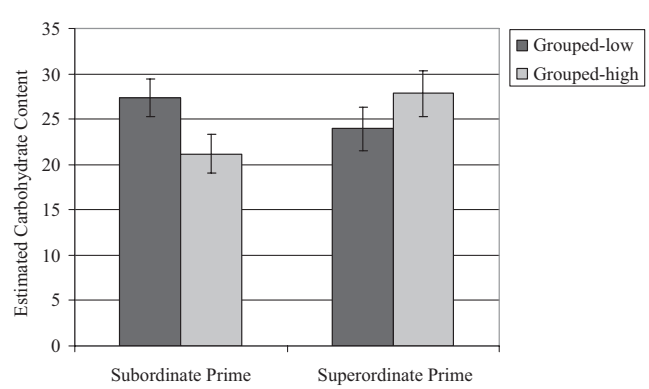

Figure 2. Evaluations from Experiment 4 of cereal carbohydrate content as a function of grouping condition and level of construal. $N = 41$.

aggregation. Instead of focusing on cues as individual pieces of information, decision-makers might recognize groups of information. We have shown how group-level and cue-level weighting strategies differ and have highlighted several factors that lead decision-makers to adopt group-level strategies. At the core of this discussion are some related predictions. First, increasing the number of grouped cues will lead decision-makers to dilute the weight placed on these cues individually. More critically, decision-makers naturally identify such groups of information or can be primed to do so.

The experiments and results presented above provide strong support for these hypotheses. In Experiments 1a and 1b, participants grouped cues spontaneously. However, the information provided across conditions was quite different, and it was therefore possible that other strategies might have produced the observed data. Experiment 2 therefore tested how different group labels might influence how people interpret otherwise identical sets of information. Although participants interpreted the data differently on the basis of how the data were grouped, it was again possible that different labels provided different information. Experiments 3 and 4 dissociated changes in content from changes in strategy by priming participants in two different ways. These experiments showed how identical sets of information could be interpreted differently depending on whether participants were in a mindset that facilitated or inhibited meaningful cue grouping.

Relation to Previous Findings

As noted in the introduction, this theory naturally extends several lines of work on how partitioning and grouping can affect probability estimation (Fischhoff et al., 1978; Fox & Clemen, 2005) and choice (Fox, Ratner, & Lieb, 2005). The present work adds to the partitioning literature by highlighting several factors which influence how people partition. Most previous work on partitioning involved experimenter-provided groups or categories, but the experiments above demonstrate how similarity, attention to dimensions for comparison, and construal level might all affect partitioning in the absence of clearly demarcated categories. Although we have focused on how people group information, these ideas might extend to how people partition events or choice options.

The experiments presented here perhaps more directly relate to a small set of experiments on how unpacking and partitioning effects carry over into cue weighting. For instance, Weber, Eisenfuhr, and von Winterfeldt (1988) asked participants to assign cue weights to attributes in a job search scenario. These attributes could either be described at a superordinate level or unpacked into smaller components. When attributes from a group were unpacked, participants ended up weighting the group more heavily overall than when the group was described at a superordinate level. For instance, the "income" attribute received more weight when it was described as the combination of "starting salary" and "future salary increases", rather than as a solitary attribute.

Martin and Norton (2009) have also recently shown how consumers can be influenced by unpacking product attributes. In one experiment, participants were asked to assign weights to a car's attributes. Some participants saw the attributes described as "practicality (safety, gas mileage, and warranty)" and "stylishness (design, stereo, horsepower)," whereas some participants saw the attributes described as "safety," "gas mileage," "warranty," and "stylishness (design, stereo, horsepower)." The latter format led to more weight being placed on "practicality" attributes than did the former, consistent with the results of Weber et al. (1988).

On the surface, these lines of research share many similarities with the current work. Yet the results from these "unpacking" studies seem to suggest that people do not naturally group information. Instead, similar cues were weighted independently. What might account for this difference?

Perhaps there was a different likelihood that people would spontaneously group the supposedly similar cues. For instance, participants might not naturally think that safety, gas mileage, and warranty are similar enough attributes to warrant grouping. Listing these components separately is therefore less like unpacking a superordinate group but, rather, more like providing participants with entirely new cues. Furthermore, asking participants to assign weights explicitly to cues makes this task less like a standard numeric judgment and more akin to the classic paradigms that have shown how unpacking influences probability assessments (Fischhoff et al., 1978).

In some ways, it may seem that cue grouping is a natural complement to these unpacking effects. If it can be shown that unpacking information changes judgments, then surely "repacking" would also change judgments. Thinking of cue grouping in this way is not inaccurate, but it does obscure an important contribution of this work.

The unpacking effects above—particularly those in the domain of cue weighting—largely provide further examples of the extensive tendency for decision-makers to use strategies similar to equal weighting at the cue level. When a cue is added to the information given, this strips some weight from *all* cues, rather than just from cues within the same group. Group-level weighting strategies demonstrate that decision-makers can break from this common tendency and focus on more nuanced relationships and similarities between pieces of information. That is, unpacking effects largely remain concerned with cue-level information integration. Cue grouping suggests that there are multiple levels at which people consider information.

Hogarth (1989) considered this possibility and questioned the factors that might lead decision-makers to consider information at the cue level or group level. Hogarth offered the possibility that expertise might moderate the level at which people weight information. This could certainly be the case, and it may be that experts adopt different mental representations which facilitate or inhibit grouping. Perhaps experts attend to different dimensions for similarity comparisons or different levels of category hierarchy.

Further Questions

The experiments described above largely focus on the factors leading to cue grouping and the consequences thereafter. This leaves unanswered any questions about how decision-makers form groups. Specifically, how do decision-makers determine whether cues are similar? Many contextual factors that affect direct similarity comparison probably also operate in the domain of cue grouping. However, there is also a factor that is slightly more specific to cue grouping, namely the target judgment being made.

Suppose that a person is asked to estimate the average temperature in the United States on a given day based on data for cities

scattered across the country. It seems natural to group cities on a north–south axis. Fargo might be grouped with Detroit, and Houston might be grouped with Little Rock. On the other hand, suppose that this person is asked to estimate the average number of hours people in the country spend watching football on television. Now Houston and Detroit might be grouped, because both cities have professional football teams, whereas Fargo and Little Rock do not. Target judgments therefore make certain dimensions salient for cue grouping. But how are these dimensions derived? Do decision-makers first think of the dimension and then search for cues that might be similar along this dimension? Or do the cues themselves also have a role in making the dimensions apparent? If a person were asked to estimate the nationwide average temperature based only on southern cities, it seems unlikely that a north–south axis would become salient without a northern city also present. Although Experiment 1b documents this effect, it does not clarify the process by which cues are grouped.

In addition to studying how groups are formed, future work should also explore how cue groups are weighted. In the experiments presented above, participants behaved as if they were equally weighting the cues within groups and equally weighting the group-level evaluations. But the specific stimuli in many of these experiments were designed to foster this hierarchical equal weighting strategy. Although this experimental approach simplified comparisons between group-level and cue-level weighting strategies, it might also obscure meaningful nuances. For instance, it is possible that decision-makers will weight particularly salient or prototypical cues more heavily. Indeed, past research suggests that this is likely (Shah & Oppenheimer, 2007, 2009).

There are also lingering questions about the processes that give rise to (or are peripherally associated with) group-level weighting. For example, do decision-makers simply allocate weights at the cue level or group level, or do they develop causal models of the relationships between cues to inform their judgments (Oppenheimer & Tenenbaum, 2010)? It is possible that decision-makers built *explanations* of why only fish carry a particular bacterium or only certain cereals have a large amount of carbohydrates. Indeed, *relevance theory* provides an outline of similar processes in the domain of induction (Medin, Coley, Storms, & Hayes, 2003). Part of this theory states that people use distinctive features of premises as the basis for induction. For example, if people were given a premise about a skunk and a premise about a zebra, then the feature "stripes" might inform their reasoning.

Relevance theory seems applicable to cue grouping as well. If people saw that fish primarily carried a certain bacterium, but that land mammals and birds did not, then they might take note of the property "lives in water" and elaborate upon the information to assume that the bacterium is water-borne. If this were the case, people might conclude that sea mammals and shore birds are also likely to carry the bacterium. The availability or salience of these unnamed groups might also influence judgments beyond the information given. In other words, decision-makers would not simply evaluate each group individually and then aggregate across the group-level evaluations. Instead, the way in which decision-makers considered information would vary based on the causal relationships that they inferred about cues and groups.

Future work might also be directed at understanding *why* people give less weight to information from the same group. It is quite possible that people believe that information from within a group is highly correlated (as it was in the studies presented here). If that is the case, then people might consciously underweight information from the same group to give more weight to uncorrelated (i.e., nonredundant) information. This would be a rather smart strategy, as judgments typically improve when people use uncorrelated information instead of correlated information (Budescu & Yu, 2007). There is some evidence that people seek out nonredundant information, but this might happen only if people have an opportunity to compare levels of cue redundancy across multiple judgment problems (Maines, 1990) or if cue redundancy is made particularly salient (Gonzalez, 1994). And there is certainly evidence that people actually prefer redundant information (Kahneman & Tversky, 1973; Slovic, 1966). It is therefore difficult to conclude whether people give less weight to grouped information because of cue redundancy, but it remains an idea worth testing.

Regardless of the underlying mechanism, the implications of cue grouping are clear. For example, consumers checking nutrition labels in the United States will see information grouped differently than in the United Kingdom. In the United States, fiber content is listed as part of the carbohydrate group, whereas in the United Kingdom, this information is listed separately. This could lead people in the United States to give less weight to fiber content in determining overall nutritional quality of an item, for its weight is diluted by the rest of the carbohydrates. Martin and Norton (2009) discussed how unpacking effects might affect cue weighting in consumer settings, and a natural extension would be to examine how cue grouping might also influence consumer behavior.

Conclusion

We began with an example of how data can be communicated differently, where different truths emerge based on how the data are grouped or aggregated. However, the larger contribution here is not that people are susceptible to such framing but, rather, that people might themselves consider multiple levels at which to weight information. Across a series of experiments, we have shown how cue grouping can introduce malleability into people's judgments. This work suggests that it is important to understand both cue-level and group-level weighting strategies. Such study brings attention to often-neglected questions of how decision-makers simplify cue weighting. And the theory presented here offers insight into what other weighting strategies might look like. It seems clear that our understanding of cue weighting will need to accommodate how people can consider pieces of information individually, as parts of groups, or possibly even at many levels.

References

Anderson, N. H. (1965). Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of Experimental Psychology, 70,* 394–400. doi:10.1037/h0022280

Budescu, D. V., & Yu, H. (2007). Aggregation of opinions based on correlated cues and advisors. *Journal of Behavioral Decision Making, 20,* 153–177. doi:10.1002/bdm.547

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34,* 571–582. doi:10.1037/0003-066X.34.7.571

Evans, J. St. B. T., Clibbens, J., Cattani, A., Harris, A., & Dennis, I. (2003). Explicit and implicit processes in multicue judgment. *Memory & Cognition, 31,* 608–618.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1978). Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 330–344. doi:10.1037/0096-1523.4.2.330

Fox, C. R., Bardolet, D., & Lieb, D. (2005). Partition dependence in decision analysis, managerial decision making, and consumer choice. In R. Zwick & A. Rapoport (Eds.), *Experimental business research* (Vol. 3, pp. 338–360). Dordrecht, the Netherlands: Kluwer.

Fox, C. R., & Clemen, R. T. (2005). Subjective probability assessment in decision analysis: Partition dependence and bias toward the ignorance prior. *Management Science, 51,* 1417–1432. doi:10.1287/mnsc.1050.0409

Fox, C. R., & Levav, J. (2004). Partition-edit-count: Naïve extensional reasoning in conditional probability judgment. *Journal of Experimental Psychology: General, 133,* 626–642. doi:10.1037/0096-3445.133.4.626

Fox, C. R., Ratner, R. K., & Lieb, D. (2005). How subjective grouping of options influences choice and allocations: Diversification bias and the phenomenon of partition dependence. *Journal of Experimental Psychology: General, 134,* 538–551. doi:10.1037/0096-3445.134.4.538

Fox, C. R., & Rottenstreich, Y. (2003). Partition priming in judgment under uncertainty. *Psychological Science, 14,* 195–200. doi:10.1111/1467-9280.02431

Fujita, K., Trope, Y., Liberman, N., & Levin-Sagi, M. (2006). Construal levels and self-control. *Journal of Personality and Social Psychology, 90,* 351–367. doi:10.1037/0022-3514.90.3.351

Gallup. (2010). *Obama weekly job approval by demographic groups.* Retrieved from http://www.gallup.com/poll/121199/Obama-Weekly-Job-Approval-Demographic-Groups.aspx

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart.* Oxford, England: Oxford University Press.

Gluck, M. A., & Bower, G. H. (1988). Evaluating an adaptive network of human learning. *Journal of Memory and Language, 27,* 166–195. doi:10.1016/0749-596X(88)90072-1

Gonzalez, R. (1994). When words speak louder than actions: Another's evaluations can appear more diagnostic than their decisions. *Organizational Behavior and Human Decision Processes, 58,* 214–245. doi:10.1006/obhd.1994.1035

Hacking, I. (2001). *An introduction to probability and inductive logic.* New York, NY: Cambridge University Press.

Hogarth, R. M. (1989). On combining diagnostic forecasts: Some thoughts and evidence. *International Journal of Forecasting, 5,* 593–597. doi:10.1016/0169-2070(89)90015-0

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237–251.

Lichtenstein, S., Earle, T. C., & Slovic, P. (1975). Cue utilization in a numerical prediction task. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 77–85. doi:10.1037/0096-1523.1.1.77

Louviere, J. J. (1984). Hierarchical information integration: A new method for the design and the analysis of complex multiattribute judgement problems. In T. Kinear (Ed.), *Advances in consumer research* (Vol. 11, pp. 148–155). Provo, UT: Association for Consumer Research.

Maines, L. A. (1990). The effect of forecast redundancy on judgments of a consensus forecast's expected accuracy. *Journal of Accounting Research, 28,* 29–47. doi:10.2307/2491245

Martin, J. M., & Norton, M. I. (2009). Shaping online consumer choice by partitioning the web. *Psychology & Marketing, 26,* 908–926. doi:10.1002/mar.20305

Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review, 10,* 517–532.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100,* 254–278. doi:10.1037/0033-295X.100.2.254

Nutrition facts [Count Chocula cereal box]. (2010). Golden Valley, MN: General Mills.

Oppenheimer, D. M., & Tenenbaum, J. B. (2010). *Categorization as causal explanation: Discounting and augmenting of concept-irrelevant features in categorization.* Manuscript submitted for publication.

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97,* 185–200. doi:10.1037/0033-295X.97.2.185

Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5,* 411–419.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker.* New York, NY: Cambridge University Press.

Permut, S. E. (1973). Cue utilization patterns in student-faculty evaluation. *Journal of Psychology: Interdisciplinary and Applied, 83,* 41–48. doi:10.1080/00223980.1973.9915589

Peterson, D. K., & Pitz, G. F. (1985). Explicit cue weighting in a prediction task. *Organizational Behavior and Human Decision Processes, 36,* 289–304. doi:10.1016/0749-5978(85)90001-9

See, K. E., Fox, C. R., & Rottenstreich, Y. (2006). Between ignorance and truth: Partition dependence and learning in judgment under uncertainty. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1385–1402. doi:10.1037/0278-7393.32.6.1385

Shah, A. K., & Oppenheimer, D. M. (2007). Easy does it: The role of fluency in cue weighting. *Judgment and Decision Making, 2,* 371–379.

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin, 134,* 207–222. doi:10.1037/0033-2909.134.2.207

Shah, A. K., & Oppenheimer, D. M. (2009). The path of least resistance: Using easy to access information. *Current directions in Psychological Science, 18,* 232–236. doi:10.1111/j.1467-8721.2009.01642.x

Slovic, P. (1966). Cue consistency and cue utilization in judgment. *American Journal of Psychology, 79,* 427–434. doi:10.2307/1420883

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance, 6,* 649–744. doi:10.1016/0030-5073(71)90033-X

Summers, S. A. (1962). The learning of responses to multiple weighted cues. *Journal of Experimental Psychology, 64,* 29–34. doi:10.1037/h0044546

Todd, F. J., & Hammond, K. R. (1965). Differential feedback in two multiple-cue probability learning tasks. *Behavioral Science, 10,* 429–435. doi:10.1002/bs.3830100406

Troutman, C. M., & Shanteau, J. (1976). Do consumers evaluate products by adding or averaging attribute information? *Journal of Consumer Research, 3,* 101–106. doi:10.1086/208657

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352. doi:10.1037/0033-295X.84.4.327

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131. doi:10.1126/science.185.4157.1124

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review, 101,* 547–567. doi:10.1037/0033-295X.101.4.547

Weber, M., Eisenfuhr, F., & von Winterfeldt, D. (1988). The effects of splitting attributes on weights in multiattribute utility measurement. *Management Science, 34,* 431–445. doi:10.1287/mnsc.34.4.431

Received May 25, 2010
Revision received September 10, 2010
Accepted September 13, 2010 ■