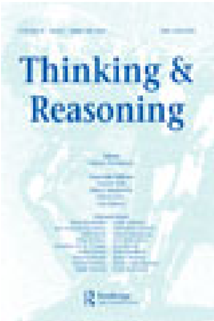

Thinking & Reasoning
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ptar20

Category-based updating

Jiaying Zhao[a] & Daniel Osherson[a]

[a] Department of Psychology, Princeton University, New Jersey, US

Published online: 24 May 2013.

Thinking & Reasoning, 2014
Vol. 20, No. 1, 1–15, http://dx.doi.org/10.1080/13546783.2013.795916

Category-based updating

Jiaying Zhao and Daniel Osherson

Department of Psychology, Princeton University, New Jersey, US

Given a prior distribution over a finite outcome space, how is the distribution updated when one outcome is excluded (i.e., assigned probability 0)? We describe two experiments in which estimated probabilities seem to "stick" to salient events. The probabilities of such events remain relatively invariant through updating. Our results suggest that the credence assigned to a salient category is sometimes more basic than the credence assigned to the constituents that comprise the category.

Keywords: Belief updating; Category; Reasoning; Probability.

Miss Marple witnessed the crime from afar but had the distinct impression that its author was a man. The only suspects were Albert, Bruce, Charles, David, Elizabeth, Florence, Gertrude, and Harriet, so Miss Marple gave higher probability to the men, resulting in the following distribution.

(1) MISS MARPLE'S PRIOR DISTRIBUTION:

Albert	Bruce	Charles	David	Elizabeth	Florence	Gertrude	Harriet
.17	.18	.17	.18	.07	.08	.07	.08

It subsequently emerged that David had a solid alibi, setting his probability to zero in Miss Marple's mind. To adjust the remaining probabilities, she reasoned as follows:

> David is not the culprit but I'm still convinced that the guilty party is a man. So I'll retain the probability of the *men* category by renormalising the probabilities of the three remaining men to add up to the same *men*'s probability as before. The probabilities of the women won't be altered.

Correspondence should be addressed to Jiaying Zhao, Department of Psychology, Green Hall, Princeton University, Princeton, NJ 08540, USA. E-mail: jiayingz@princeton.edu

We thank Derek Shiller for a lecture on probability that inspired the work reported here. We also thank Vincenzo Crupi, Konstantinos Hadjichristidis, David Over, Katya Tentori, and Steven Sloman for helpful comments.

Specifically, for each of Albert, Bruce, and Charles, Miss Marple multiplied the prior probability by

$$\frac{Pr(\text{Albert}) + Pr(\text{Bruce}) + Pr(\text{Charles}) + Pr(\text{David})}{Pr(\text{Albert}) + Pr(\text{Bruce}) + Pr(\text{Charles})}$$

to obtain:

(2) Miss Marple's posterior distribution:

Albert	Bruce	Charles	David	Elizabeth	Florence	Gertrude	Harriet
.2288	.2424	.2288	0	.07	.08	.07	.08

We see that the probabilities of *men* and *women* in (2) are .7 and .3, respectively, just as they were in (1).

Miss Marple's reasoning is compelling. David's exoneration does not alter her impression of having seen a man commit the crime; it just makes it more likely that one of the other men is guilty. Indeed, if Bruce and Charles came up with alibis as well, that would go to show that Albert is likely the culprit! It must be admitted, however, that a posterior distribution different from (2) results from thinking along Bayesian lines. A Bayesian would renormalise all seven remaining probabilities in light of David's zero. Specifically, each of the seven probabilities would be divided by their sum to reach:

(3) The Bayesian posterior distribution:

Albert	Bruce	Charles	David	Elizabeth	Florence	Gertrude	Harriet
.2073	.2194	.2073	0	.0854	.0976	.0854	.0976

The choice between (2) and (3) reflects different strategies for attaching credence to events. The Bayesian proceeds by distributing probabilities over outcomes in a sample space; an event inherits its probability from the sum of the probabilities of the outcomes that comprise it (Hacking, 2001). A different tradition sorts the available evidence by the specific event to which it is relevant, allowing an event to be directly supported even if its constituent outcomes are not (Shafer, 1976). Miss Marple's reasoning seems more congenial to the latter approach inasmuch as she conserves her credence in *men* despite David's alibi undermining one of its elements.

It is not our present goal to judge Miss Marple.[1] Indeed, both her strategy and that of the Bayesian may prove useful in different contexts (Shafer & Tversky, 1986). We wish merely to identify some circumstances in which

[1] She is the amateur detective in Agatha Christie's crime novels.

Marple-type reasoning (rather than Bayesian) seems to represent common intuition. For this purpose it will be helpful to introduce some conventions.

The experiments reported below elicit distributions over eight-member sample spaces. Each space decomposes into four-member salient categories analogous to *men* and *women* for Miss Marple. Typically, one of the categories attracts higher prior probability than its complement. In fact, we will only consider trials in which all the probabilities assigned to the high-probability category exceed all the probabilities assigned to the low-probability category. Such a prior distribution will be called *acceptable*, and is illustrated by (1). After an acceptable prior is established, we set one of its eight outcomes to zero probability, and invite the participant to update. The zero can be chosen from the high-probability or low-probability category. In either case, *category-based updating* consists of replacing each of the three remaining probabilities in that category with the result of multiplying it by the ratio

(4) $$\frac{\text{sum of the four prior probabilities in the category}}{\text{sum of the three prior probabilities that were not set to zero in that category}}.$$

The probabilities in the complementary category are untouched. Category-based updating is illustrated in (2). By *Bayesian updating* we mean the usual operation, illustrated in (3).

For each acceptable prior distribution produced by an experimental participant, we ask whether the participant's posterior is numerically closer to the category-based versus the Bayesian update. It will be seen that category-based is closer but only if the outcome that is given zero probability is drawn from a high-probability category; otherwise, Bayesian updating tends to better approximate the participant's posterior.

There are, of course, alternative ways to implement category-based updating. Instead of multiplying by (4) to preserve the ratios of the probabilities for Albert, Bruce, and Charles, Miss Marple could have distributed David's defunct probability equally among the latter three suspects yielding the posterior:

(5) MISS MARPLE'S POSTERIOR DISTRIBUTION, IF SHE USES IMAGING:

Albert	Bruce	Charles	David	Elizabeth	Florence	Gertrude	Harriet
.23	.24	.23	0	.07	.08	.07	.08

This update can be understood as applying Lewis's 1986 idea of *imaging* inasmuch as David's probability is distributed to the three suspects most similar to him in virtue of common gender.[2] Comparison of (5) and (2)

[2] We are indebted to David Over for this observation.

reveals that the two strategies for category-based updating are numerically very close when the members of a given category start off with similar prior probabilities. Such is the case for the judgements collected in the experiments described below. The important point is the psychological naturalness of category-based updating, whatever the particular scheme for implementing it. We focus on (4) because it presents the simplest contrast with Bayesian updating; whereas the Bayesian preserves proportions of all non-excluded events, Miss Marple preserves only proportions of same-category events. As noted above, the latter scheme is descriptively superior to Bayesian updating (for major categories). The superiority is preserved by use of imaging, but we leave the matter implicit for brevity. For more on imaging in human reasoning, see Baratgin and Politzer (2010).

Probability updating has long been of interest to psychologists (Edwards, 1954; for an overview, see Oaksford & Chater, 2007). Inspiring our Marple vignette, Robinson and Hastie (1985) and Van Wallendael and Hastie (1990) examine updating using crime mysteries and find that participants often ignore the requirement that posterior distributions sum to unity. This issue does not arise for the experiments reported here inasmuch as the computerised interface that collects judgements forces renormalisation.[3] The present focus is rather on the "locality" of updating, that is, whether trimming the sample space affects all remaining members or just those that resemble the zeroed element. So far as we know, this topic has not yet been directly broached. After presenting data from two experiments, we consider what the phenomenon of category-based updating reveals about human credence.

EXPERIMENT 1

Method

Forty-one Princeton undergraduates participated in exchange for course credit. The experiment was built around the eight objects shown at the top of Table 1 below. Participants were first familiarised with the objects, then completed eight trials. Each trial had two parts, yielding prior and posterior distributions.

Creating a subjective prior distribution. For each trial, 100 objects were randomly drawn with replacement according to the trial's probability distribution, shown in the corresponding row of Table 1 (each row represents one trial). Each drawn object was presented for 100 ms at the centre of a

[3] In any event, pilot testing revealed that Princeton undergraduates recognise that both prior and posterior distributions must sum to unity.

TABLE 1
Stimuli and objective distributions for Experiment 1

Major category									Excluded item
blue	0.20	0.20	0.20	0.20	0.05	0.05	0.05	0.05	2
	0.20	0.20	0.20	0.20	0.05	0.05	0.05	0.05	6
green	0.05	0.05	0.05	0.05	0.20	0.20	0.20	0.20	7
	0.05	0.05	0.05	0.05	0.20	0.20	0.20	0.20	3
triangle	0.20	0.20	0.05	0.05	0.20	0.20	0.05	0.05	5
	0.20	0.20	0.05	0.05	0.20	0.20	0.05	0.05	3
pentagon	0.05	0.05	0.20	0.20	0.05	0.05	0.20	0.20	4
	0.05	0.05	0.20	0.20	0.05	0.05	0.20	0.20	2

Each row corresponds to one trial, and shows the probability of a given object being selected in a single draw (there were 100 draws per trial). The leftmost column indicates the major category for each trial. The rightmost column shows which object was revealed *not* to be the one selected in the covert 101st draw. The objects are numbered from left to right.

computer screen, followed by a 50 ms blank interval. The rapid sampling prevented counting the occurrences of the eight objects. After the 100 draws were completed, the participant was told that a further 101st random draw had been completed covertly. For each of the eight objects, the participant was asked to estimate its probability of being the 101st draw. These eight estimates (one for each object) constitute the *prior distribution* for the trial in question. Thus, each participant created his or her own prior distribution, typically distinct from that of any other participant.

Creating a subjective posterior distribution. To complete the trial, the participant was next instructed that a certain object was *not* the one sampled in the 101st draw. The right-hand column of Table 1 shows which object was excluded for each trial. The participant then re-estimated the probability of each of the seven objects being the 101st draw. Probability zero was filled in for the excluded object; the participant was asked to produce the remaining seven estimates. The resulting probabilities (including the zero) constitute the *posterior distribution*.

When estimating both the prior and posterior distributions, participants were allowed to proceed only if their estimates summed to unity. The eight trials were administered to participants in individually random order.

Results

In a given trial, a four-member subset of the eight objects is called *major* just in case:

(a) the members of the subset are either of the same colour (all blue or all green), of the same shape (all triangles or all pentagons), or of the same line texture (solid or dotted);[4] and
(b) the prior probabilities assigned to members of the subset are each greater than all prior probabilities assigned to the remaining four objects.

The complement of a major category is called *minor*. It is easy to verify that given a prior distribution, there is at most one way to divide the eight objects into major and minor categories. A trial will be called *acceptable* if it produced major/minor categories, and none of the prior probabilities were taken to be zero. All other trials were dropped from further analysis. An acceptable trial in which the excluded object (i.e., the one set to zero by the experimenter) falls into the major category will be called *a major trial*. The remaining trials (in which the excluded object falls into the minor category) are called *minor trials*. (In our story, Miss Marple is confronted with a major trial based on the category *man*.) The 41 participants produced 131 major trials and 124 minor trials.

Each acceptable trial is associated with a *Bayesian* and with a *category-based* update. The Bayesian update is the result of dividing each of the seven non-zeroed probabilities in the prior distribution by their sum. The category-based update is the result of multiplying each of the three non-zeroed prior probabilities from the category that holds the zeroed object by the ratio defined in (4); the probabilities of the category without the zeroed object are not changed. It may be helpful to illustrate Bayesian and category-based posterior distributions starting from the distribution shown in the first row of Table 1 (a major trial). (As indicated earlier, this distribution need not have served as any participant's prior.) If the probability of the second object is set to zero, then the rival posterior distributions are as follows.

Bayesian	.25	0	.25	.25	.0625	.0625	.0625	.0625
Category-based	.2666	0	.2666	.2666	.05	.05	.05	.05

[4] Major categories based on line texture were rare because the objective probabilities never favoured such a grouping.

Likewise, for the second row of Table 1 (a minor trial), the two distributions are:

Bayesian	.2105	.2105	.2105	.2105	.0526	0	.0526	.0526
Category-based	.2	.2	.2	.2	.0666	0	.0666	.0666

Of course, the Bayesian and category-based updates were always computed from the participant's own prior distribution.

For each acceptable trial, the Bayesian and category-based updates were compared with the participant's posterior distribution. Specifically, we computed the mean absolute deviation between the eight numbers of the participant's posterior versus the Bayesian update, which is referred to as the *Bayesian predictive error*. Likewise, we computed the mean absolute deviation between the eight numbers of the participant's posterior versus the category-based update, which is referred to as the *category-based predictive error*. We now present statistics about these two kinds of predictive errors. Major and minor trials are discussed separately.

Major trials. Of the 131 major trials produced in the experiment, 91 yielded posterior distributions in which the category-based predictive error was smaller than the Bayesian predictive error. The average category-based predictive error for major trials was .165 ($SD = .12$) whereas the average Bayesian predictive error was .181 ($SD = .10$). For each participant, we counted the number of major trials in which the category-based distribution had a predictive error that was smaller than the Bayesian distribution did. For 29 out of 41 participants, there were more major trials in which the category-based predictive error was smaller than the Bayesian predictive error. The reverse was true for 7 participants, and 5 participants were tied (including 1 participant who produced no major trials). Finally, for each participant, we computed the average category-based predictive error as well as the average Bayesian predictive error on major trials (the participant without major trials was excluded). Across participants, the mean average category-based and Bayesian predictive errors were .167 ($SD = .074$) and .184 ($SD = .065$), respectively. They were reliably different via a paired t-test [$t(39) = 2.58$, $p = .01$].

Minor trials. Of the 124 minor trials produced in the experiment, 78 yielded posterior distributions in which the Bayesian predictive error was smaller than the category-based predictive error. The average category-based predictive error for minor trials was .094 ($SD = .057$) whereas the average Bayesian predictive error was .082 ($SD = .041$). For each participant, we counted the number of minor trials in which the category-based distribution had a predictive error that was smaller than the Bayesian distribution did.

For 11 of 41 participants, there were more minor trials in which the category-based predictive error was smaller than the Bayesian predictive error. The reverse was true for 23 participants, and 7 participants were tied (including 1 participant who produced no minor trials). Finally, for each participant, we computed the average category-based predictive error as well as the average Bayesian predictive error on minor trials (the participant without minor trials was excluded). Across participants, the mean average category-based and Bayesian predictive errors were .094 ($SD = .036$) and .085 ($SD = .031$), respectively. A paired t-test produced a trend for smaller Bayesian predictive errors [$t(39) = 1.71, p = .10$].

Additional analyses. We examined how many objects changed probability in the transition from prior to posterior distributions. The object assigned zero probability is not included in this count, leaving just the seven others. Within major trials, 73.4% of the three non-zeroed objects in the major category shifted probability whereas only 25% of the four objects in the minor category shifted probability. The difference was reliable via a paired t-test [$t(39) = 8.62, p < .001$]. Within minor trials, 24.7% of the three non-zeroed objects in the minor category shifted probability whereas 27.6% of the four objects in the major category shifted probability. The difference was not reliable [$t(39) = 0.55$, $p = .58$]. These numbers represent the grand mean of averages for individual participants. Thus, major trials provoked more asymmetry than minor trials in the number of objects whose probabilities changed from prior to posterior. We also found that in 57 of the 131 major trials (43%), the sum of the major category probabilities remained the same between the prior and posterior distributions. The same is true for 33 of the 124 minor trials (26.6%). These results highlight the tendency in major trials to preserve the probabilities accruing to each of the two categories.

Overall, Experiment 1 reveals a reliable tendency to preserve the probability of a major category when one of its members is excluded. In contrast, when a member of a minor category is excluded, the update shows a trend toward the Bayesian solution. In both cases participants likely compromised between category-based versus Bayesian updating rather than purely siding with either. For major categories, the relative attraction is most clearly reflected in the number of trials in which category-based updating leads to a posterior distribution that is closer to participants' judgements than is the Bayesian posterior (91 versus 40). The probabilities appearing in minor categories are smaller overall (by definition) but would still show a category-based bias if this policy were more attractive than the Bayesian one. The absence of such a tendency (46 favouring category-based updating compared to 78 for Bayesian) suggests a strategic difference in updating for the two situations.

EXPERIMENT 2

In Experiment 1, the objective probabilities governing the sampling of objects were designed to encourage acceptable priors in the minds of participants. Experiment 2 relied on the participants' background knowledge for the same purpose.

Method

Thirty undergraduates (19 female, mean age 21.2 years, $SD = 1.1$) from Princeton University participated in exchange for course credit. Participants completed ten trials. Each trial had two parts, yielding prior and posterior distributions.

Creating a subjective prior distribution. In a given trial, eight familiar items were presented; the task was to assign each item its (subjective) probability of exceeding the other items along a certain criterion. For example, one trial presented four different headphones and four different GPS devices; for each item, participants gave their probability that it was the most expensive among the eight according to Amazon.com. Another trial presented four foreign (non-USA) cities and four USA cities; in this case, participants stated their probability that each city had the highest population among the eight. The ten trials are summarised in Table 2 below. The eight items of a given trial were presented simultaneously via computer monitor in individually randomised position. As with Experiment 1, each participant was free to estimate his or her own prior distribution. We relied on a pilot study ($N = 10$) to verify that Princeton undergraduates perceived the categories as intended. All participants divided the eight stimuli into the two categories indicated in Table 2. The only exception was the trial involving cars, which elicited little respect for the division of budget versus luxury models; we nonetheless retained the car trial in the experiment.

Creating a subjective posterior distribution. In the second part of a trial the participant was informed that a certain item did not, in fact, exceed the others along the criterion for that trial. For example, they were informed that a particular GPS device was *not* the most expensive item among the eight indicated in the first row of Table 2. Participants then re-estimated the probabilities for the remaining seven items in the trial. These seven probabilities (plus zero for the excluded item) constitute the subjective posterior distribution. The last column in Table 2 shows which item was excluded in a given trial.

For both distributions, the participant was allowed to proceed only if her or his estimates summed to unity. The ten trials were administered to participants in individually random order.

TABLE 2
Stimuli used in Experiment 2

Criterion	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Excluded item
Highest price	headphone	headphone	headphone	headphone	GPS	GPS	GPS	GPS	6
Highest price	TV	TV	TV	TV	air cond.	air cond.	air cond.	air cond.	3
Highest price	budget car	budget car	budget car	budget car	luxury car	luxury car	luxury car	luxury car	7
Largest student body	ivy-league	ivy-league	ivy-league	ivy-league	state univ.	state univ.	state univ.	state univ.	8
Largest population	Seoul	Mumbai	Tokyo	Jakarta	New York	Los Angeles	Chicago	Houston	5
Greatest life expectancy	Malaysia	Singapore	S. Korea	China	Algeria	Turkey	Egypt	Morocco	2
Highest GDP	India	Indonesia	Taiwan	S. Korea	Sweden	Belgium	Switzerland	Holland	6
Highest birth rate	Niger	Zambia	Uganda	Kenya	Malaysia	Philippines	Vietnam	Fiji	6
Highest profit	Exxon Mobil	Apple	Chevron	Citigroup	Shell	Gazprom	PetroChina	Samsung	2
Largest USA box office	Avatar	Titanic	Star Wars	Dark Knight	Toy Story	Finding Nemo	Lion King	Shrek	8

Each row corresponds to one trial, and consists of eight items. The leftmost column indicates the criterion for which items in a trial were judged. The rightmost column shows which item was excluded. To illustrate, the first row is a trial in which four different brands of headphones and four different brands of GPS devices were displayed to participants via screenshots. For each item, participants gave their probability that it was the most expensive of the eight (according to Amazon.com). Subsequently, the participant was informed that the sixth item (the second GPS) was not the most expensive, and they re-estimated probabilities for the remaining seven items. In a given trial, intuition points to a division between the first four items and the last four.

Results

As noted above, in each trial the eight items may be intuitively divided into two sets of four. In trial 1, for example, the two sets are the four headphones versus the four GPS devices. Table 2 shows the divisions used. In a given trial, we qualify as *major* either of these intuitive subsets provided that each of the probabilities assigned to its members exceed each of the probabilities assigned to the members of its complement. The complement of a major subset is called *minor*. As with Experiment 1, a trial will be called *acceptable* if it produced major/minor categories, and none of the prior probabilities were taken to be zero. All other trials were dropped from further analysis. An acceptable trial in which the excluded item falls into the major category will be called a *major trial*. The remaining trials (in which the excluded item falls into the minor category) are called *minor trials*. The 30 participants produced 90 major trials and 60 minor trials.

Each acceptable trial is associated with both a Bayesian and a category-based update as explained above. For each acceptable trial, we computed the mean absolute deviation between the eight numbers of the participant's posterior versus the Bayesian update (the *Bayesian predictive error*). Likewise, we computed the mean absolute deviation between the eight numbers of the participant's posterior versus the category-based update (the *category-based predictive error*). We now present statistics about these two kinds of predictive errors for major and minor trials separately.

Major trials. Of the 90 major trials produced in the experiment, 64 yielded posterior distributions in which the category-based predictive error was smaller than the Bayesian predictive error. The average category-based predictive error for the major trials was .135 ($SD = .114$) whereas the average Bayesian predictive error was .168 ($SD = .080$). For each participant, we counted the number of major trials in which the category-based distribution had a predictive error that was smaller than the Bayesian distribution did. For 20 out of 30 participants, there were more major trials in which the category-based predictive error was smaller than the Bayesian predictive error. The reverse was true for 4 participants, and 6 participants were tied. Finally, for each participant, we computed the average category-based predictive error as well as the average Bayesian predictive error on major trials. Across participants, the mean average category-based and Bayesian predictive errors were .133 ($SD = .103$) and .171 ($SD = .068$), respectively. They were reliably different via a paired *t*-test [$t(29) = 3.47, p = .002$].

Minor trials. Of the 60 minor trials produced in the experiment, 42 yielded posterior distributions in which the Bayesian predictive error was smaller than the category-based predictive error. The average category-based predictive error for minor trials was .119 ($SD = .093$) whereas the average Bayesian predictive error was .100 ($SD = .068$). For each participant, we counted the number of minor trials in which the category-based distribution had a predictive error that was smaller than the Bayesian distribution did. For 7 of the participants, there were more minor trials in which the category-based predictive error was smaller than the Bayesian predictive error. The reverse was true for 17 participants, and 6 participants were tied (including 2 who produced no minor trials). Finally, for each of the 28 participants who produced minor trials, we computed the average category-based predictive error as well as the average Bayesian predictive error on minor trials. Across participants, the mean average category-based and Bayesian predictive errors were .119 ($SD = .093$) and .100 ($SD = .068$), respectively. A paired t-test produced a trend for smaller Bayesian predictive errors [$t(27) = 1.160$, $p = .256$].

Experiment 2 thus produced results which are similar to those of Experiment 1. Category-based updating prevailed when a member of a major category was excluded whereas there was a non-significant tendency to update in the Bayesian way for minor categories.

DISCUSSION

Both experiments suggest that people often assign credence to complex events (i.e., categories) in a more fundamental way than they do to the event's atomic constituents, as changes in the status of an event's constituent do not always propagate to the event itself. Thus, when the presentation of objects in Experiment 1 imparted high probability to the last draw being blue (for example), learning that a specific blue object was not drawn had relatively little impact on the belief in blue. Likewise, in Experiment 2, if a participant attached high probability to a European nation having the highest GDP among the eight listed then they tended to retain that conviction even if one of the European nations was excluded as having the highest GDP.

Such stability points to the use of "basic belief assignments" (Shafer, 1976; Smets & Kennes, 1994) when evaluating events defined over a finite sample space, rather than standard probability distributions. Given an event E made up of elementary outcomes $e_1 \cdots e_n$, it may happen that belief in E is psychologically prior to belief in the e_i. In this case, credence flows "down" to the e_i from E, although there might also be a contribution from the e_i that prevents them from having uniform probability. This is different from the standard picture of credence flowing "up" to E from the conviction first garnered by each e_i. In a very helpful discussion, Kotzen (2012) puts the matter as follows (adapting to the present context). Credence flows down if the agent's *reason* for her belief in the e_i is her belief in E. Conversely, credence flows up if her reason for believing E is based on her beliefs about the e_i. Thus, in Experiment 1, the projection of many blue objects directly supported this category, and its probability flowed down to the more specific blue objects like the blue, solid triangle. The latter objects were apparently registered less directly. In Experiment 2, the GPS devices seemed generically more expensive than the headphones, and sent their credence down to specific GPS models.

On the other hand, our results suggest that low-probability events fail to provide reason to believe in their constituents. Indeed, minor categories gave no evidence of category-based updating, conforming instead to the Bayesian rule. The role of degree of belief in the choice of update might be clarified in future studies by posing direct questions about the probabilities of categories. Other subtleties affecting the subjective probabilities of categories are discussed in Lagnado and Shanks (2003), Murphy, Chen, and Ross (2012), and Tversky and Koehler (1994). Their relation to category-based updating (and updating more generally) remains to be explored.

Consider again a major category C in our experiments. When the probability of an elementary event in C is set to zero, its prior probability p must be shifted to other events. We have focused on just one alternative to Bayesian updating, namely, dividing p among the three remaining members of C in a proportional way, as described by (4). Alternatives come to mind, notably, dividing p into three equal parts and adding a part to each of the three remaining probabilities in C (as noted earlier, this scheme resembles "imaging"; Lewis, 1986). For our data, the additive and proportionate rules deliver very similar numbers. In the more general case, however, the additive rule exhibits counter-intuitive behaviour if one member of C has very low probability; updating additively can impart a posterior that is too high (Kotzen, 2012 offers a compelling example along with an additional objection to the additive rule). But it remains possible that some alternative to our version of category-based updating would improve the prediction of posterior probability.

We conclude with two observations. Fix an event space, and consider an agent whose beliefs are represented by a probability distribution Pr_1. Suppose that an event B is learned, and consider the agent's new distribution Pr_2. If the agent is Bayesian then for every event A, $Pr_2(A) = Pr_1(A|B)$. The latter equality represents the *invariance* of the agent's beliefs (through the experience of learning B) about probabilities with conditioning event B. Our results suggest that for major categories, participants engaged in category-based updating, rather than Bayesian updating, which implies that they violated invariance. For example, the prior distribution (1) yields $Pr_1(\text{Albert}|\text{not-David}) = .2073$, which is not equal to $Pr_2(\text{Albert}|\text{not-David}) = .2288$ in the posterior (2). Since we did not collect estimates of $Pr_1(A|B)$ before the participant learned which object was excluded, $Pr_1(A|B)$ could not be directly observed, but only inferred from the prior distribution. Future studies can request estimates of $Pr_1(A|B)$ and $Pr_2(A|B)$ in order to directly assess invariance in category-based updating. For more discussion of the role of invariance in reasoning, see Over and Hadjichristidis (2009), and Zhao and Osherson (2010).

The final observation concerns the successive elimination of elements from a major category. In Miss Marple's case, suppose that David comes up with an alibi, then Charles does the same, followed by Bruce. Miss Marple might continue to believe that the culprit was a man, and hence ultimately focus her credence on Albert. Alternatively, Miss Marple might give up the belief that the culprit was a man. Future work is needed to examine the conditions under which the major category is maintained or abandoned.

Manuscript received 9 November 2012
Revised manuscript received 5 April 2013
First published online 24 May 2013

REFERENCES

Baratgin, J., & Politzer, G. (2010). Updating: A psychologically basic situation of probability revision. *Thinking & Reasoning*, *16*, 253–287.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, *51*, 380–417.

Hacking, I. (2001). *An Introduction to Probability and Inductive Logic*. Cambridge: Cambridge University Press.

Kotzen, M. (2012). *Where should displaced credence go?* (Unpublished paper). University of North Carolina, US.

Lagnado, D. A., & Shanks, D. R. (2003). The influence of hierarchy on probability judgment. *Cognition*, *89*, 157–178.

Lewis, D. (1986). Probabilities of conditionals and conditional probabilities II. *The Philosophical Review*, *5*, 581–589.

Murphy, G. L., Chen, S. Y., & Ross, B. H. (2012). Reasoning with uncertain categories. *Thinking & Reasoning*, *18*, 81–117.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: Oxford University Press.

Over, D., & Hadjichristidis, C. (2009). Uncertain premises and Jeffrey's rule. *Behavioral and Brain Sciences*, *32*, 97–98.

Robinson, L. B., & Hastie, R. (1985). Revision of beliefs when a hypothesis is eliminated from consideration. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 443–456.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton University Press.

Shafer, G., & Tversky, A. (1986). Languages and designs for probability judgment. *Cognitive Science*, *9*, 309–339.

Smets, P., & Kennes, R. (1994). The Transferable Belief Model. *Artificial Intelligence*, *66*, 191–243.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, *101*, 547–567.
Van Wallendael, L. R., & Hastie, R. (1990). Tracing the footsteps of Sherlock Holmes: Cognitive representations of hypothesis testing. *Memory & Cognition*, *18*, 240–250.
Zhao, J., & Osherson, D. (2010). Updating beliefs in light of uncertain evidence: Descriptive assessment of Jeffrey's rule. *Thinking & Reasoning*, *16*, 288–307.