Knowing Too Much: Expertise Induced False Recall Effects in Product Comparison

RAVI MEHTA

JOANDREA HOEGG

AMITAV CHAKRAVARTI *

* Ravi Mehta (ravi.mehta@sauder.ubc.ca) is assistant professor of business administration at the University of Illinois at Urbana-Champaign, 350 Wohlers Hall, Champaign, IL, 61820. JoAndrea Hoegg (joey.hoegg@sauder.ubc.ca) is assistant professor of marketing at Sauder School of Business, University of British Columbia, 2053 Main Mall, Vancouver, BC, Canada V6T 1Z2. Amitav Chakravarti (achakrav@stern.nyu.edu) is associate professor of marketing at Leonard N. Stern School of Business, New York University, 40 West 4th Street, New York, NY 10012. The authors wish to acknowledge the helpful input of the editor, associate editor and reviewers, and to thank Joseph W. Alba, Darren W. Dahl, Chris Janiszewski, and Tom Meyvis for their comments on earlier drafts of the manuscript.

ABSTRACT

A long history of research has shown that experts' well developed knowledge structures provide numerous advantages in memory-based decisions and tasks. More recently, research has shown that in certain situations experts' more detailed knowledge can hinder memory performance by resulting in the creation of false memories. We add to this growing research by showing how experts can fall prey to a different type of false memory when making product comparisons. Four studies demonstrate that in a product comparison context, in their attempt to make options more comparable experts inadvertently "fill in the gap" by aligning nonalignable features in memory. This results in the false recall of aligned features that did not appear in the original descriptions. We identify experts' higher sense of accountability for their judgments, coupled with their highly developed schemas, as the mechanism underlying the effect. Research has traditionally shown that experts' well developed knowledge structures provide numerous advantages in memory-based decisions and tasks (Alba and Hutchinson 1987; Chase and Simon 1973; Voss, Vesonder, and Spilich, 1980). Recent work, however, has demonstrated that expertise can at times be a double-edged sword (Baird 2003; Castel et al. 2007; Marchant et al. 1991). Experts have been found to be at a disadvantage when they are unable to make use of knowledge that would normally help them solve a problem, or when task demands run counter to highly proceduralized behaviors (Marchant et al. 1991). Their well developed schemas have also been shown to interfere during information retrieval, leading to false recall of information not previously presented (Baird 2003; Castel et al. 2007). In the current work we extend prior research by demonstrating expert fallibility, both in terms of recall and decision making, in the context of memory-based product comparisons. In doing so, we reveal a different kind of false recall that stems directly from the product comparison context.

Specifically, we suggest that when engaging in memory-based product judgments, consumer experts will attempt to make extensive comparisons (Alba and Hutchinson 1987) in order to arrive at a justifiable decision. To illustrate, imagine a consumer attempting to recall features from different video game systems when trying to make a choice from memory. For a given pair of brands under consideration, say an Xbox 360 and a Playstation 3, it is quite likely that some attributes will be directly comparable across the two products while some will not. That is, attribute comparison will involve both aligned differences, which are feature values provided for both product options and hence easily comparable on that dimension (e.g., 60 GB hard drive for the Xbox versus a 120 GB hard drive for the Playstation) and non-aligned differences, which are feature values described only for one of the product options (e.g., BluRay video playback listed for the Playstation and no information about video playback provided for

the Xbox) (Zhang and Markman 1998). We suggest that while making memory-based comparisons experts attempt to maximize the comparability of options, and in doing so, they inadvertently align the non-aligned differences across products. This attempt to maximize comparability may result in false recalls as experts inadvertently "fill in the gaps" and recall a feature value for the option that did not have the feature or feature value listed in the original description. Depending on the feature values falsely recalled, this may, in turn, decrease product evaluation accuracy.

Why would this effect occur for experts but not for novices? We offer a two-part explanation. First, consumer experts have a far better developed schema than novices; thus, consistent with prior research on expert false memories, their more complex schemas increase the likelihood of falsely recalling an associative link from memory. However, this fact alone explains only why consumer experts recall more information, some of it false, not why the product comparison context has a particular influence.

The second element, we argue, stems from a heightened sense of accountability that experts naturally feel, relative to novices (King and Summers 1970; Leonard-Barton 1985; Simonson and Nye 1992). This heightened sense of accountability leads experts to consider more attributes and attempt to conduct a more detailed comparative assessment in their product evaluations (Tetlock 1986). This effort will, in turn, cause them to align the nonalignable differences by "filling in" an attribute value that was not originally presented. Novices, by comparison, do not feel this heightened sense of accountability and hence attend primarily to the easier-to-recall, already alignable differences (Markman and Medin 1995; Zhang and Markman 1998, 2001). In essence, we argue that a combination of experts' well-developed schemas and the heightened sense of accountability that they naturally feel, leads to suboptimal feature recall

4

and product evaluation.

Across four studies we investigate experts' false recalls in a product comparison context and extend the current understanding of how expertise influences memory-based product comparison. Our findings show that experts are prone to false recalls that occur as a result of an attempt to align non-aligned differences, which in turn contributes to sub-optimal product evaluations. We also reveal the underlying process of expert accountability that induces this type of false memory. Finally, we demonstrate conditions under which expert advantages can be restored. Below we provide the theoretical basis for our work, drawing from research in expertise, memory-based product comparison, and accountability. We then present a series of studies that test our predictions. We conclude with a discussion of the implications of our work and highlight directions for future research.

THEORETICAL DEVELOPMENT

The Role of Expertise in Product Judgments

In marketing environments, memory tasks often involve consideration of brand options, where product attributes need to be recalled and compared (Biehal and Chakravarti 1986; Kardes, Muthukrishnan, and Pashkevich 2004; Lynch and Srull 1982). For example, when considering options at different stores, recall of attribute details from alternatives at previous stores is critical to reaching a decision. Past research in preference formation has found that in order to arrive at a preference, consumers often represent the attributes of options and compare them (Bettman, Johnson and Payne 1991; Simonson and Tversky 1992). Indeed, some researchers contend that in consumption contexts comparisons are readily generated (Wang and Wyer 2002; but see Sanbonmatsu, Kardes, and Gibson 1991), as evidenced by consumers' tendency to spontaneously place features of one option in correspondence with features of another (Markman and Medin 1995; Mather et al. 2005).

When attempting to recall feature information in order to make product judgments, conventional wisdom would suggest that expertise in a product category is an asset. An extensive body of literature supports this common belief, providing evidence that expertise leads to superior performance in a variety of contexts, including product information search, product comprehension, and feature memory (Brucks 1985; Cowley and Mitchell 2003; Huffman and Houston 1993; Roehm and Sternthal 2001). Expert consumers have complex, highly developed schemas of their subject area, which tends to make them more effective at learning and remembering information (Cowley and Mitchell 2005).

There is, however, some evidence, albeit not from the consumption domain, that these well-formed schemas could also be a detriment. When attempting to remember a list of items, for example, experts' schemas can interfere with their attempts to recall the items (Baird 2003; Castel et al. 2007). In the domain of investments, although experts, compared to novices, recalled more words correctly from an investment-related list, they were also more likely to falsely recall investment-related words that were not presented (Baird 2003). In other words, the strength of associative links in an expert's memory can increase correct and false recalls. As experts attempt to recall a presented item, a non-presented associate can also be spontaneously activated, leading to high levels of both correct *and* false recalls (Underwood 1965).

Expert memory failure in a product comparison context. The attempt to maximize comparability has been shown to be a general tendency for all consumers, regardless of expertise

(Sjoberg 1972). However, we believe that varying levels of domain knowledge may lead to differences in how comparability is maximized. Specifically, when comparing products with aligned differences, creating correspondence is straightforward and will occur similarly for experts and novices. However, when products also have non-aligned differences, experts will be more likely than novices to attempt to align the non-aligned differences. As a consequence, when recalling a feature for one option, experts may unknowingly falsely recall a comparable counterpart for the other option (Mather et al. 2005). For example, a consumer shopping for video game systems may read information about the features of a Playstation 3 and an Xbox 360 and later attempt to recall the features of both brands. If information about the processing speed of the system was presented for the Playstation but not for the Xbox, when later recalling this information, the consumer may erroneously recall a processing speed value for the Xbox.

The type of recall error just described stems directly from experts' efforts to make options comparable in memory; hence, we refer to this type of false recall as an *attempt-to-align false recall*. It is important to note that these recall errors differ from those studied in the list recall paradigm (Baird 2003, Castel et al. 2007, Roediger and McDermott 1995). In such research, participants memorize a list of words (e.g., night, dream, bed) related to a target word (e.g., sleep) that is omitted from the presented list. The target word is later incorrectly recalled as having been present because of its association to the list words. Our paradigm, on the other hand, examines false memories created in an attempt to make nonalignable alternatives more comparable. That is, the focus is on maximizing comparability of the two options.

We propose that when attempting to recall features of product options, experts' attempts to conduct a detailed comparison of all available attributes will lead them to attempt to align the non-aligned differences. These attempts will, in turn, result in the creation of attempt-to-align false recalls. More formally,

H1: When attempting to recall features from two options in a comparison paradigm, expert consumers will commit more attempt-to-align false recalls than will novices.

This assertion naturally begets the question of why we expect these false recalls to occur for experts but not for novices. As noted earlier, we believe that a combination of experts' welldeveloped knowledge structures and the heightened sense of accountability that they naturally feel, leads to these attempt-to-align false recalls. We elaborate on this causal chain next.

The Role of Accountability and Knowledge Structures in Attempt-to-Align False Recalls

Experts are, by definition, expected to know more about a given domain than others. Research has demonstrated that experts tend to provide more detail than others when making judgments (Black et al. 1998; Evans et al. 2007) and tend to use a greater number of criteria in making decisions (Ballou 2001). In comparison contexts it has been shown, in diverse domains, that the expert narrative commonly relies on extensive comparisons to make its case (e.g., Coley et al. 2005; Ballou 2001; Evans et al. 2007; Mackiewicz 2010). In a study of online product reviews, for example, Mackiewicz (2010) found that one of the most prominent manners in which reviewers assert their expertise is by drawing comparisons between the reviewed product and related and relevant products. Indeed, that experts are expected to provide as much detailed information as possible is a belief so well-ingrained, that even when lay people are asked to assume the role of experts and dispense advice, they use more pieces of information than they would have had they been deciding for themselves (Kray and Gonzales 1995). We suggest that one reason for this tendency to provide more thorough and detailed information is that experts naturally feel a heightened sense of accountability for their judgments. Because they are looked to as sources of information and are expected to be able to provide detailed information about products (Leonard-Barton 1985; Simonson and Nye 1992), self-described experts feel that they should be able to prove that they are indeed capable of providing reliable information. This higher sense of accountability and consequent concern with making judgments that others would approve of (Kardes et al. 2004) may lead them to consider more information, attend to more features, and attempt to draw comparisons (Tetlock 1983; Tetlock and Boettger 1989). Formally,

H2: Experts feel a higher sense of accountability for their responses than novices.

Accountability, knowledge structures, and attribute alignability in product comparison. We have argued above that experts' feelings of accountability lead them to conduct a more detailed comparison of all available attributes. It is this greater effort to compare individual features across options that, we believe, leads experts to commit more recall errors than novices. However, it is still unclear why experts would commit attempt-to-align errors in particular, and not just any recall error. To understand why experts are more likely to commit attempt-to-align errors in particular, it is important to consider the joint role of experts' sense of accountability, their well-developed knowledge structures, and the differential ease-of-recall of alignable vs. nonalignble attributes.

As noted earlier, when evaluating products consumers attempt to compare options by placing features of one option in correspondence with those of the other (Bettman, Johnson, and Payne 1991; Simonson and Tversky 1992). All consumers, experts and novices alike, are likely to recall and compare the easily retrievable, alignable features that have been presented for both

product options (Markman and Medin 1995; Markman and Loewenstein 2010; Zhang and Markman 1998, 2001). However, experts' heightened sense of accountability will make them go beyond the alignables, and attempt to recall the more difficult-to-retrieve nonalignable features. We propose that rather than improving expert accuracy, this added effort to retrieve and compare the nonalignables results in more attempt-to-align false recalls. Experts' knowledge structures further contribute to these false recalls since their richly-developed schemas enable them to more easily, albeit inaccurately, recall a comparable attribute.

Unlike experts, novice consumers do not naturally feel this sense of accountability, and therefore will generally restrict their retrieval efforts to the easy-to-recall, alignable features (Markman and Medin 1995; Zhang and Markman 1998, 2001). Even if novices were somehow made to feel more accountable and encouraged to retrieve the nonalignable features, their inferior schemas would make it less likely that they would be able to generate alignable features to enable comparison. Thus, we identify experts' higher sense of accountability, coupled with their highly developed schemas, as the mechanism underlying the effect noted under H1.

Importantly, feelings of accountability should not alter experts' tendency to make errors in recalling the values of aligned features. Incorrect recall of the value of a feature that was already aligned does not help to make the options more comparable. Rather, such recalls are more akin to false recalls examined in list paradigms (e.g., Baird 2003; Roediger and McDermott 1995), which stem solely from an expert's greater familiarity with the product category. Feelings of accountability should also not alter experts' tendency to falsely recall a feature for a single option that had not been previously presented. Erroneously remembering that one product had a particular feature but not recalling anything about that feature for the other option indicates no attempt to make the options comparable, but rather reflects merely the mistaken recall of a feature for one of the options. Because we argue that accountability leads experts to try to increase the comparability of the two options, we expect that these types of errors would not be driven by accountability. Only recalls specifically stemming from efforts to align the attributes (i.e., attempt-to-align false recalls) should be influenced by feelings of accountability. Thus, we hypothesize the following:

H3A: Accountability mediates the relationship between expertise and attempt-to-align false recalls.

If accountability encourages experts to make the product options more comparable, resulting in attempt-to-align false recalls, releasing experts from their feelings of accountability should mitigate such errors. In contrast, releasing novices from any feelings of accountability should have no effect on their attempt-to-align false recalls. More formally,

H3B: When expert consumers are made to feel unaccountable for their responses, attemptto-align false recalls are reduced.

Outcome versus Process Accountability

We have argued that feelings of accountability lead experts to make more attempt-toalign false recalls. However, given that accountability is often shown to debias consumers and improve decision making, it is useful to understand why in memory-based product comparison, accountability would lead to memory detriment. That is, what is it about feelings of accountability that underlies the recall difficulty? We suspect that in this type of memory task, feelings of accountability cause experts to focus more on the steps involved in making their judgments, rather than the final decision itself, and ironically this focus leads to memory errors. We borrow from the literature on process and outcome accountability to support our proposition.

Prior research has demonstrated that feelings of accountability can be related to a focus on either the *outcome* of a decision or the *process* by which the decision is reached (Lerner and Tetlock 1999). In terms of judgment accuracy, past research has shown that a process focus leads to greater accuracy (Simonson and Staw 1992). Outcome-accountable participants focus entirely on reaching the right answer and thus simply attempt multiple paths to reach it without attending to the process involved in each path. In contrast, by focusing on the decision steps, processaccountable participants generally improve both in accuracy and in the correspondence between their accuracy and their confidence (Siegel-Jacobs and Yates 1996; Simonson and Staw 1992).

However, for memory-based product comparison tasks that we are examining, we expect the opposite result. Process accountability leads individuals to "engage in more evenhanded evaluation of alternatives" and consideration of more features (Lerner and Tetlock 1999, p. 258). The fact that consumer experts naturally feel the need to provide detailed information, often comparing and contrasting the features of the available options, suggests to us that experts tend to feel accountable for the process by which they reached their decision. This leads them to attend to all product features, both aligned and non-aligned. Consequently, when retrieving product features, they attempt to align the non-aligned features, resulting in attempt-to-align false recalls and suboptimal evaluations. In short, we suspect that process accountability is what drives experts' attempt-to-align false recalls. Thus, in contrast to prior research showing an advantage for process accountability, for experts a focus on the process may be detrimental.

Outcome accountability, on the other hand, may actually mitigate experts' susceptibility to the false memory effect, since it encourages individuals to focus only on finding the right

answer and not on the reasons behind their decisions. As consumers gain expertise, they learn to simplify tasks, essentially automating parts of the process (Alba and Hutchinson 1987; Stenberg 1997). This can allow experts to easily reach a correct outcome despite being unable to articulate how or why they reached it (Langer and Imber 1979; Wimmers et al. 2005). Removing the expectation of having to prove how they reached a solution frees experts, allowing them to focus only on finding the right answer, a task for which expertise is helpful.

In summary, we propose that when experts engage in memory-based product comparison, the sense of accountability they experience leads them to focus on the process of their decision, rather than on the outcome. This stems from a natural feeling that they should be able to prove how they made their judgments. This focus on the decision process impairs experts' memory and evaluations. By making experts focus only on outcome and hence releasing them from the burden of explaining the process details, recall accuracy and evaluation should improve. In contrast, for novices, we predict the reverse. That is, consistent with research demonstrating the superiority of process accountability, encouraging novices to focus on the decision process is likely to improve both their memory of features and the quality of their judgments.

H4A: As expertise increases, a focus on the decision outcome rather than the decision process, will reduce attempt-to-align false recalls.

H4B: As expertise increases, a focus on the decision outcome rather than the decision process, will improve the quality of product evaluations.

We tested our hypotheses through four experiments. Experiment 1 investigated whether experts commit more attempt-to-align false recalls than novices when engaging in memorybased product comparison (hypothesis 1). In addition, experiment 1 examined whether experts naturally feel a higher sense of accountability than novices when engaging in memory-based product comparison (hypothesis 2) and whether this feeling is responsible for the increase in attempt-to-align false recalls (hypothesis 3A). Experiment 2 examined whether reducing feelings of accountability could debias experts and improve their memory performance (hypothesis 3B). In experiments 3 and 4 we looked at the potentially detrimental influence of attempt-to-align false recalls on product evaluation. Experiment 4 also delved further into the role of accountability in expert memory performance by examining the different effects of process versus outcome accountability. We tested whether focusing experts on accountability for the outcome of their decision rather than the process used to reach it would mitigate attempt-to-align errors and improve evaluations (hypotheses 4A and 4B).

EXPERIMENT 1

Experiment 1 was conducted to test Hypotheses 1, 2, and 3A. We expected that when attempting to recall features from a previously presented two-option choice set, expert consumers would generate not only more correct recalls but also more false recalls. More importantly, in line with our key research hypothesis (H1), we predicted that experts would generate a higher proportion of attempt-to-align false recalls as compared to novices.

The study was also designed to test whether experts naturally feel a greater sense of accountability for their responses than novices (H2); and whether the presence of this selfimposed accountability would lead to a higher proportion of attempt-to-align false recalls, but not other types of false recalls (H3A). To these ends, we administered individual-level measures of accountability to all participants and tested for mediation.

Method

Stimuli. We created two video game consoles as stimuli. We first consulted with a professional video game designer to generate a list of important video game console features. The goal was to select appropriate features that could be used to create two brand options. Once the features were determined, the designer assisted in splitting them into configurations for the two brands such that the options had eight features each and were equivalent in terms of feature attractiveness (see appendix A). For each option, half of the features were aligned, i.e., they were comparable with features of the other option (e.g., for one option "No hard drive" and for the other "20Gb hard drive"). The other half of the features for each option were non-aligned, i.e., there was information about the feature for one of the options but not for the other option (e.g., for one option "Bluetooth wireless networking" whereas for the other option there was no mention of wireless networking). To counterbalance which features were aligned and which were not, we used four versions of the two-option choice set. Each feature appeared as an aligned feature in one version and as a non-aligned feature in another version. The option features were listed in random order, so that the four aligned features were not in the same location in the list as their corresponding features in the other option (see appendix B).

Expertise. As product experience, which is a source of consumer expertise, is more strongly related to subjective judgments of knowledge than to objective knowledge (Park, Mothersbaugh, and Feick 1994), we utilized subjective knowledge as our measure of expertise. Participants were asked four questions related to their experience with video game consoles, each on 7-point scale. The items included, "How would you rate yourself as a gaming enthusiast," "How often do you play gaming systems like Playstation and Xbox," "In general how much do you know about video game consoles like Playstation and Xbox (not the games but gaming systems)," and "How would you rate your level of knowledge on video game consoles relative to your peers." The responses to the four items were averaged to create an expertise index.

Accountability. To measure feelings of accountability, all participants answered accountability-related questions. In line with previous literature suggesting that the idea of being judged or getting evaluated leads to a higher level of felt accountability (Lee et al. 1999; Lerner and Tetlock 1999; Tetlock and Boettger 1989), we adopted the following four questions to measure felt accountability: 1) "I felt accountable for the choices I made", 2) "I was concerned about the possibility of making a mistake", 3) "It was important to me to get things right" and 4) "I felt accountable for how I evaluated the options." Each item was rated on a 7-point scale with 'not at all' and 'very much' as the low and high anchors, respectively.

Procedure. One hundred thirteen undergraduate students (sixty-three women) from a large North American university participated in the experiment. Participants were invited to a computer lab in groups of five to ten students per session, and were seated at computer stations. Instructions on the screen explained that they would be presented with descriptions of various options from a product category and would be asked to evaluate the brands. They were then presented with one of the four versions of the stimuli, containing two options of the video game consoles, Option A and Option B, and were asked to examine the options carefully. The feature information remained on the screen for two minutes, after which the computer program automatically proceeded to the next screen. Participants then answered the four expertise-related questions and completed an unrelated filler task for 20 minutes. After completion of the filler task they were asked to recall the features of the two options they had previously examined. Participants were presented with two text boxes, side-by-side on the screen, in which they noted

down the features they could recall for Option A and Option B respectively. The accountability measures were taken upon completion of the recall task. At the end of the study participants provided some demographic information and were dismissed.

Analysis

Recall coding. Across all studies, two independent coders who were blind to the study conditions individually coded the recall responses according to the criteria described below and as outlined in table 1. Agreement between the two coders was 98% for the current study and between 97% and 100% for subsequent studies. Disagreements were resolved through discussion. Recalls were broadly coded into the number of features correctly recalled and the number of features falsely recalled. Correct recalls were then subdivided into recall of aligned features and non-aligned features. Although not a central focus of this research, we included correct recalls in the current study to show replication of prior research on expert memory.

The falsely recalled features were subdivided into three categories. Category 1 comprised the primary focus of our investigation, attempt-to-align false recalls. These were false recalls made in an effort to align the product features across options to maximize comparability. Two types of false recalls were considered attempt-to-align false recalls. First, if a participant mistakenly recalled a feature for one of the options, such that it matched a non-aligned feature that was presented for the other option (e.g. recalling "1080p video resolution" for option B when no video resolution information was presented for that option but "720p video resolution" information was presented for option A) this was coded as an attempt-to-align false recall. In addition, if a participant mistakenly recalled a new feature that had not been presented for either option, and provided a value for both product options (e.g., "No option to play a DVD" for option A and a matching value, "Option to play DVD" for option B when nothing was mentioned about DVDs for either option), this was also considered an attempt to align false recall. In both types of false recalls, the memory error enabled comparison between the two options by creating aligned features; thus, they were coded as attempt-to-align false recalls.

Category 2 false recalls occurred when a feature that was mistakenly recalled did *not* increase the comparability of the two options. We refer to these errors as incorrect value recalls. This category was primarily comprised of errors made in recalling specific values of previously presented features. For example, when a participant recalled a feature that had been previously presented for both options but recalled the incorrect value for one or both of the original options (e.g., recall of 40 GB hard drive for Option A and 20 GB for Option B, when it was actually 60 GB for A and 20 for B), this fell into category 2. Similarly, recall of the incorrect value for a feature that was presented for only one of the options also constituted a Category 2 recall. In addition, there were a few recalls of a feature for one of the options (but not both) when the feature had not previously been presented for either option. Although these comprised a very small proportion of the total recalls, they are false recalls that do not increase the comparability of the options, and as such fell into Category 2.

Category 3 recalls occurred when participants recalled a feature that had been presented for both options, but switched the values between the options. For example, recalling that Option A had a 40 GB hard drive and Option B had a 60 GB hard drive when they were actually the reverse. Because the correct values were actually recalled but simply for the wrong options, these may reflect confusion as to the source of the value rather than a truly false recall; hence we refer to them as confusions (Burke and Srull 1988). Although they are not central to our hypotheses we include them for completeness.

In addition to the three categories described above, there were a few random comments by participants that did not comprise any kind of recall and were removed from the analysis. These were generally comments about the study, such as "I hate video games" or "This is boring." These types of comments accounted for 2% to 5% of the total recalls across all studies.

Insert table 1 about here

Expertise. Participants' expertise was assessed through responses to the video game console knowledge questions. Responses loaded on the same factor and were averaged to form an expertise index ($\alpha = .94$). For all analyses, expertise was treated as a continuous variable.

Results

We begin by presenting the results which are central to our research hypotheses, i.e., the false recalls. For each category of false recall we calculate the proportion of false recalls relative to the total number of recalls made. Because experts are expected to have a higher number of total recalls, they are more likely to have a higher absolute number of errors as well. Hence, in order to examine the effect of expertise, it is the proportion of false recalls relative to total recalls rather than the absolute number of false recalls that is of interest. We also report the correct recalls and total recalls to provide a full picture of the recall pattern. For illustration purposes, appendix C provides a summary of incidence rates for all types of recalls, reported at one standard deviation above and below average expertise.

Category 1: Attempt-to-align false recalls. These false recalls were those which enabled comparison between the options. Supporting H1, regression analysis indicated that increased expertise led to the generation of a higher proportion of attempt-to-align false recalls (β = .25, *t* = 2.70, *p* < .01).

Category 2: Incorrect value recalls. Consistent with previous literature, as expertise increased, so did the proportion of the false recall of incorrect values or attributes, i.e., recalls that were unrelated to making the options more comparable, ($\beta = .18$, t = 1.92, p < .06).

Category 3: Confusions. Expertise had no effect on recalls for which participants reversed the values for an aligned feature ($\beta = -.12$, t = -1.24, ns).

Total false recalls. Total false recalls was the sum of Category 1, 2 and 3 false recalls. As expertise increased, the proportion of total false recalls also increased ($\beta = .16$, t = 1.75, p = .08).

As noted we analyzed the different types of false recalls as a proportion of total recalls rather than the absolute number of errors because we believed that the error rates were more informative than the absolute numbers. However, an alternative view is that regardless of the number of overall recalls, experts should also commit fewer absolute errors than novices, and hence the absolute numbers would also be of interest. Therefore, we also analyzed the recall data utilizing the absolute number of errors under each category. Mirroring the results for proportions, experts committed a higher number of total false recalls ($\beta = .25$, t = 2.77, p < .05), attempt-toalign false recalls ($\beta = .25$, t = 2.72, p < .01) and incorrect value recalls ($\beta = .22$, t = 2.32, p <.05). However, no effect of expertise was observed for the number of confusions ($\beta = -.032$, t <1). Across all our studies, the pattern and significance of results for the absolute number of false recalls was consistent; thus, for brevity, for all subsequent analysis we report only the proportion results. *Correct recalls.* Because the proportion of correct recalls is equivalent to one minus the proportion of incorrect recalls, and as such relatively uninformative, we report the results for correct recalls using the absolute numbers. The regression analysis indicated that increased expertise led to a higher number of correct recalls ($\beta = .24$, t = 2.57, p < .05). Breaking this into correct recall of non-aligned and correct recall of aligned features, we observed that non-aligned features recalled increased significantly with expertise ($\beta = .30$, t = 3.27, p < .01), but the slope of the regression line was not significant for aligned features ($\beta = .09$, t < 1).

Total recalls. As there is no proportion to report for total recalls, we also conducted the analysis using the absolute numbers. Not surprisingly, we found a significant positive relationship between expertise and total recalls ($\beta = .33$, t = 3.62, p < .001).

Accountability. Responses to the four accountability items were averaged to create an accountability index ($\alpha = .81$), such that a higher score on the accountability index indicated higher level of felt accountability. Supporting hypothesis 2, we observed that as expertise increased so did the sense of accountability ($\beta = .26$, t = 2.78, p < .01).

Mediation analysis. Following procedures recommended by MacKinnon, Lockwood, and Williams (2004), we tested our hypotheses using a bootstrapping procedure in which the significance of the indirect (i.e., mediation) effects was tested. In the bootstrapping test developed by Preacher and Hayes (2008), the 95% bias-corrected bootstrap confidence intervals were obtained using 5000 bootstrap samples. Supporting H3A, the results show that the indirect effects (i.e., the effect of expertise on attempt-to-align false recalls through accountability) were significant. The upper and lower confidence intervals did not include zero (CI = .0004 to .0060) also indicating that the indirect effects were significant.

As expected, no mediation effect of accountability was observed on the relationship

between expertise and Category 2 false recalls, nor on total false recalls. Similarly, no mediation was observed on the relationship between expertise and correct recalls or total recalls.

Gender. There was a significant main effect of gender on expertise, with males being higher in expertise than females ($M_{Males} = 4.43$ vs. $M_{Females} = 2.23$, F(1, 111) = 77.77, p < .001). Males did have higher correct recall than females (F(1, 111) = 6.45, p < .05). However, gender had no effect on commission of attempt-to-align false recalls or incorrect value recalls (both p's > .20). Moreover, gender did not interact with expertise for any of our key variables of interest (all p's > .20); thus it is not discussed further.

Discussion

The results of experiment 1 provided initial support for hypotheses 1, 2, and 3A. We showed that in a memory-based product comparison context, experts were prone to a different kind of false recall than those previously investigated, one that stemmed from attempts to make the options more comparable: attempt-to-align false recalls. We also demonstrated that experts experience a higher sense of accountability than novices and that this elevated sense of accountability mediates the relationship between expertise and attempt-to-align false recalls.

Consistent with previous research using single lists (e.g. Baird 2003; Castel et al. 2007), we also found that experts make more correct recalls, although this was driven by nonalignable recalls. This result is not surprising, as prior research has demonstrated that alignable features are easier to recall than nonalignable (Markman and Loewenstein 2010; Zhang and Markman 2001). Both experts and novices are likely to attend to alignable features and hence remember them equally well. It is only the more difficult nonalignables for which experts are likely to have higher correct recall.

Also in line with prior research, experts committed more incorrect recalls overall, and were more prone to recalls of incorrect values of aligned features. Importantly, however, and consistent with our theorizing, we found that these other types of recall errors were not mediated by feelings of accountability.

We observed no effect of expertise on confusions. It may be that expertise truly has no effect on of the incidence of confusion; however, there were very few confusion errors making it impossible to draw firm conclusions.

Although our experiment employed video games, which are a male-dominated product, the results of our study indicated that gender was not driving our effects. Indeed we found no effects of gender on any of our key variables. This result is consistent with other research on videogames, which has demonstrated that video games affect male and female participants similarly, albeit the size of the effect differs by gender (Anderson and Bushman 2001; Carnagey and Anderson 2005; Feng, Spence, and Pratt 2007).

Experiment 1 showed that accountability drives experts' attempt-to-align false recalls. If this is the case, then relieving experts of their feelings of accountability should reduce these types of false recalls, but have little impact on experts' other false recalls. We tested this hypothesis in the next experiment. Experiment 2 was also designed to examine the role of involvement in the relationship between expertise and false recalls. Although, previous research has argued that the two constructs are distinct (Lee et al. 1999), in the next study we empirically tested whether this bears out in our product comparison context. For that purpose, we recorded participants' involvement in the study.

EXPERIMENT 2

Method

Stimuli. Experiment 2 utilized the videogame console stimuli used in the previous experiment. A 3 (Accountability: control vs. unaccountable vs. accountable) x expertise (continuous variable) design was employed in this experiment. Expertise was measured as in the previous experiment ($\alpha = .94$). Accountability was manipulated in line with previous research (Tetlock and Boettger 1989). At the beginning of the experiment, all participants were told that they would be presented with two options from a product category and would be asked to evaluate the brands. Those in the "accountable" condition were then told, "Part of today's study is on video games and is being conducted in conjunction with a local video game company. At the end of the study the researcher may ask you to participate in a follow-up interview with a representative of the company to understand exactly why you made the choices you made." Participants in the "unaccountable" condition were told, "Part of today's study is on video games and is being conducted in conjunction with a local video game company. However, please note that all information you provide during the following study will be completely confidential and cannot be connected to you personally." For control condition participants, no additional instructions were provided.

To investigate the role of involvement, all participants answered four involvement questions at the end of the study. The items, measured on a 7-point scale from 1 = "not at all" to 7 = "very much" were, 1) How much did you enjoy this study, 2) How motivated were you in completing this study, 3) How much effort did you spend in completing this study, and 4) How

interesting do you think this study was.

Procedure. One hundred forty eight undergraduate students completed the study in return for partial course credit. They were invited to the computer lab in groups of ten to fifteen and were randomly assigned to one of the three accountability conditions. Participants were first given the accountability instructions followed by the presentation of the videogame stimuli. The study then proceeded in the same manner as Experiment 1. Finally, all participants answered four items measuring the level of involvement.

Analysis

Recalls were coded following the procedure described in experiment 1. Data were then analyzed in accordance with the Aiken and West (1991) approach. Responses on the expertise index were mean centered and treated as a continuous variable in the model. As the accountability factor had three levels, two dummy variables were created with the 'accountable' condition as the reference category. That is, for dummy 1, the unaccountable condition was coded as 1 while the accountable and control conditions were coded as 0. For dummy 2, the control condition was coded as 1 while accountable and unaccountable conditions were coded as 0. Next, we created two-way interaction terms involving expertise and the two dummy variables. We then regressed each type of recall on expertise, accountability dummy 1, accountability dummy 2 and the interaction terms.

Results

Category 1: Attempt-to-align false recalls. As hypothesized, a significant two-way interaction of expertise and accountability dummy 1 (i.e., accountable-unaccountable) emerged for the proportion of attempt-to-align false recalls generated ($\beta = -.34$, t = -3.23, p < .01). However, the interaction between expertise and accountability dummy 2 (i.e., accountablecontrol) was non-significant (t < 1). To further examine the two-way interactions we plotted the graphs at one standard deviation above and below the mean of the expertise scale (see figure 1). Consistent with H3B, at higher expertise, individuals generated a significantly smaller proportion of false recalls in the unaccountable condition (M = 1.89%) than in the accountable condition (M= 10.52%; β = -8.63, t = -2.96, p < .01). However, at lower expertise no difference was observed between the total number of false recalls generated in the unaccountable (M = 5.60%) and accountable conditions (M = 1.20%; $\beta = 4.40$, t = 1.47, ns). Contrast analysis indicated a nonsignificant difference in the proportion of false recalls generated between the control condition and the accountable condition both when expertise was low ($M_{control} = 1.65\%$ vs. $M_{accountable} =$ 1.20%; $\beta = .45$, t < 1) and when it was high ($M_{control} = 7.12\%$ vs. $M_{accountable} = 10.52\%$; $\beta = -3.40$, t = -1.23, ns).

Further, consistent with our theorizing, analyses indicated that a higher proportion of attempt-to-align false recalls were generated at higher levels of expertise as compared to lower levels of expertise in the accountable condition ($M_{high} = 10.52\%$ vs. $M_{low} = 1.20\%$; $\beta = 2.96$, t = 3.19, p < .01) and marginally more in the control condition ($M_{high} = 7.12\%$ vs. $M_{low} = 1.65\%$; $\beta = 1.53$, t = 1.53, p = .10). However no difference was observed between attempt-to-align false recalls generated at higher levels versus lower levels of expertise in the unaccountable condition ($M_{high} = 1.89\%$ vs. $M_{low} = 5.60\%$; $\beta = -1.17$, t = -1.32, ns).

Insert figure 1 about here

Category 2: Incorrect value recalls. As expected, no significant two-way interactions were observed between expertise and either of the accountability dummy variables for false recalls unrelated to attempts to increase the comparability of the options (accountable-

unaccountable dummy: $\beta = -.06$, t < 1; and accountable-control dummy: $\beta = -.50$, t = 1.06, *ns*).

Category 3: Confusions. The analysis revealed no interaction between expertise and the accountable-unaccountable dummy variables for confusions ($\beta = -.24$, t < 1). Additionally, although not central to our theorizing, there was a marginally significant interaction between expertise and the accountable-control dummy ($\beta = .17, t = 1.72, p = .09$). Total and correct recalls. No significant interaction between expertise and either of the accountability dummy variables was observed for either total recalls or correct recalls (all t's< 1). Also, no significant main effect of either of the dummy variables emerged for total recalls or correct recalls. We did observe a significant main effect of expertise for total recalls made ($\beta = .15$, t = 3.08, p < .01). There was no main effect of expertise on correct recall, although the effect was directionally consistent with experiment 1 ($\beta = .08$, t = 1.45, p < .15). Although not a central hypothesis for us, these results near statistical significance, and their pattern is consistent with the standard expectation that experts would generate more correct recalls than novices. Further, considering correct recalls in terms of aligned and nonaligned features we again found that there was no effect of expertise on aligned features ($\beta = .03, t < 1$), but there was a marginally significant effect of expertise on nonaligned features ($\beta = .10, t = 1.70, p < .10$).

Involvement. Responses to the four involvement items were averaged to create an

involvement index ($\alpha = .78$). No significant two-way interactions were observed between expertise and either of the accountability dummy variables (accountable-unaccountable dummy: $\beta = -.03$, t = -.93, ns; and accountable-control dummy: $\beta = -.01$, t = -.33, ns) for level of involvement. However, as conjectured, a significant main effect of expertise emerged, such that as expertise increased so did involvement ($\beta = .10$, t = 3.30, p < .001). Hence, the results suggest that regardless of feelings of accountability, higher expertise leads to higher involvement. We also ran another planned contrast and found a significant positive effect of expertise on involvement for each of the accountability conditions (accountable: $\beta = .28$, t = 3.33, p < .01; unaccountable: $\beta = .24$, t = 2.24, p < .05; control: $\beta = .17$, t = 2.04, p < .05).

Discussion

The results from experiment 2 provided additional support that experts' feelings of accountability when making product comparisons render them more prone to attempt-to-align false recalls. When experts were explicitly relieved of accountability, we observed a significant reduction in their attempt-to-align false recalls. Removing accountability made experts essentially equivalent to novices in the proportion of attempt-to-align false recalls generated. In contrast, attempt-to-align false recalls were significantly higher for experts relative to novices when respondents were specifically made accountable or when no instructions regarding accountability were given. No significant difference was observed between these two conditions, supporting our premise that experts' higher sense of accountability is naturally experienced.

Importantly, the results also demonstrated that experts' feelings of accountability did not alter the commission of false recalls unrelated to attempts to increase option comparability. In all three accountability conditions, experts simply made more incorrect value recalls, consistent with prior research. This finding provided support for our proposition that attempt-to-align false recalls and incorrect value false recalls stem from different processes.

Experiment 2 also considered the role of involvement. We found that making experts unaccountable for their judgments did not reduce their level of involvement, supporting the notion that the two constructs are distinct (Lee et al. 1999). We observed a reduction in attempt-to-align recalls when accountability was reduced but involvement remained high. Hence, it appears unlikely that involvement is a key driver of attempt-to-align false recalls.

So far, we have focused only on the effect of expertise on the generation of false recalls in memory-based product comparison. However, from a practical perspective, the question is whether these types of false recalls have any effect on product evaluation. The answer is not obvious. Previous research in information recall suggests that experts can be worse at recalling information about how a decision was reached, but still make better decisions (Wimmers et al. 2005). For example, although doctors are worse than third year interns at recalling the exact information they used to make a diagnosis, their diagnoses are more accurate (Patel and Groen 1991; Schmidt and Boshuizen 1993).

We expect that in a product comparison context, false recalls, specifically attempt-toalign false recalls, have the potential to adversely affect decisions. Because experts make more effort to compare all features in order to make a judgment, they pay relatively more attention to non-aligned or unique features (Houston, Sherman, and Baker 1991), which results in a more complex judgment task. Efforts to align the options should simplify evaluation by making the options more comparable. However, because the features are falsely recalled, they may bias judgment by inadvertently reducing the perceived difference, i.e. the spread between the alternatives (Liberman and Forster 2006).

Certainly in some cases, falsely recalled attributes could actually aid judgment if they happen to emphasize that a superior product was indeed superior and simply helped to spread the alternatives further and ease the judgment task (Liberman and Forster 2006). However, in cases where the unique features do not already clearly suggest that one product is superior to another, the falsely recalled information will likely be detrimental or at least unhelpful. If such false recalls are relied upon, they could hinder the comparison process, resulting in suboptimal decisions for experts relative to novices. We tested this possibility in the next experiment.

EXPERIMENT 3

Method

Stimuli and design. As we were interested in measuring the accuracy of product evaluations, we altered our stimuli for this experiment, again with the assistance of a professional video game designer. First, in addition to the four aligned and four non-aligned features, we added four common features, i.e., features that were identical for both options (Zhang and Markman 2001). Hence each of the options had 12 features instead of 8 as in previous experiments, which improved the realism of the task. Second, we reallocated the features to both options so as to make option B superior to option A (see appendix D).

To test that option B was indeed perceived superior to option A by our participant population, we conducted a pretest with 44 people from the same population as our study participants. We presented them with the 12 stimuli features in random order and asked them to rank the features from 1 (most important) to 12 (least important). We then averaged the ranks for each type of feature creating average ranks for the aligned features, the non-aligned features of option A and the non-aligned features of option B. Paired *t*-tests among three feature sets indicated that the features on which option B was better (i.e., the aligned features) had a significantly higher average ranking (M = 5.78) than the non-aligned features of option A (M = 6.59; t = 2.24, p < .05) and the non-aligned features of option B (M = 7.12; t = 3.36, p < .01). No significant difference between the average rankings of non-aligned features of option A and option B was observed (t = 1.35, ns); thus, overall, option B was considered better on the more important features. The stimuli were then compiled into two counterbalanced versions (i.e., in terms of which option was option A and which was B). Within each version, the order of the features was randomly determined.

The experiment tested whether expertise would impact the proportion of attempt-to-align false recalls, and whether these recalls would lead to impaired product evaluation. Expertise was measured as in previous studies. Product evaluation was recorded at the end of the study. Participants evaluated each of the two options on two 7-point scales: Bad – Good and Low Quality – High Quality.

Procedure. Fifty-two undergraduate students participated in this study in exchange for partial course credit. All participants were presented with the videogame stimuli for three minutes, responded to the expertise related questions, completed a series of filler tasks, and then recalled the product option features. After the recall task participants evaluated each of the two options on the two 7-point scales. One individual indicated participating in one of the earlier studies and was hence removed from the analysis.

Results

False recalls. Consistent with the previous studies, experts generated significantly higher proportion of attempt-to-align (Category 1) false recalls ($\beta = .27$, t = 1.98, p = .05). In addition, increased expertise led to a marginally significantly higher proportion of incorrect value (Category 2) recalls ($\beta = .26$, t = 1.93, p = .06). However, there was no effect of expertise on confusions (Category 3) (t < 1).

Evaluation. Responses to both evaluation items were averaged to create an index for option A ($\alpha = .85$) and for option B ($\alpha = .80$). To create an overall evaluation score, the evaluation index of option A was subtracted from the evaluation index for option B, so that higher numbers indicated higher evaluations for the objectively superior option B. To examine the effect of expertise on product evaluation, we regressed the evaluation index on the expertise scale. Results indicated that increased expertise significantly decreased the evaluation of objectively superior option B ($\beta = -.30$, t = -2.19, p < .05).

Mediation analysis. Next we tested whether attempt-to-align false recalls or incorrect value recalls mediated option evaluation. Using the bootstrap method we found evidence that attempt-to-align false recalls mediated evaluation. The upper and lower limits of 95% confidence intervals did not include zero (CI = -.64 to -.05), indicating that the indirect effects were significant. No mediation effect was observed for incorrect value recalls.

Correct recalls. The increase in correct recalls was directionally consistent with prior studies, but only marginally significant ($\beta = .25$, t = 1.82, p = .07). Also, as expertise increased the non-aligned features recalled increased marginally ($\beta = .24$, t = 1.73, p = .09), but the slope of the regression line was not significant for correctly recalled aligned features ($\beta = .19$, t = 1.37, *ns*). Also, no difference was observed for common features ($\beta = .12$, t = .87, *ns*).

Total recalls. Replicating the results from previous studies, we observed a significant positive relationship between expertise and total recalls, such that as expertise increased, so did the number of total recalls ($\beta = .43$, t = 3.32, p < .01).

Mediation analyses showed no mediating effects for either total recalls or correct recalls.

Discussion

The results of experiment 3 provided evidence that attempt-to-align false recalls can lead to lower decision quality in memory-based product comparisons. Although option B was both designed (by a video game expert) and pretested to be superior to option A, participants with greater expertise viewed the options as more similar. Moreover, this result was mediated by attempt-to-align false recalls. It appears that the attempt-to-align false recalls may have biased judgments by altering the influence of the nonalignable features. Interestingly, and as implied by previous research on information and list recall, incorrect value recalls did not mediate the relationship between expertise and judgment. This null result is not unexpected as prior research has demonstrated that all consumers, novices and experts alike, will use alignable attributes in judgments (Markman and Loewenstein 2010).

It seems that the bias in judgment occurs as an unfortunate consequence of experts' efforts to improve the comparability of the options and simplify the decision task.

In our final experiment, we return to the issue of expert accountability to gain insight into the role that accountability plays in experts' attempt-to-align false recalls. Specifically, study 4 examines precisely how the feeling of accountability differentially impacts experts in memorybased product comparisons. Prior research has shown that accountability leads to more careful consideration of all pieces of information (Tetlock and Boettger 1989). The finding that attemptto-align false recalls occur because of experts' feelings of accountability supports the possibility that experts may be focusing too much on their judgment processes rather than trying to reach a correct judgment outcome. If so, then shifting experts' focus to the outcome should reduce attempt-to-align false recalls (H4A). This should also lead to improvement in evaluation quality for experts (H4B).

EXPERIMENT 4

Method

Design and stimuli. In this experiment we tested whether making consumers accountable for the outcome of a decision or for the process by which they reached it altered their tendency to generate attempt-to-align false recalls. Thus the experiment employed a 2 (Accountability: outcome vs. process) x Expertise (continuous) between subjects design. We utilized the same stimuli employed in experiment 3. Expertise was measured as in prior experiments ($\alpha = .88$).

We manipulated outcome and process accountability by providing different instructions to participants depending on condition (Simonson and Staw 1992). Outcome-accountable participants were told that they would be asked to justify only their final choice and no questions would be asked about the process by which they reached their decision. Process-accountable participants were told that they would be asked to justify their decision process and no questions would be asked about their final judgment. Option evaluation was measured as in experiment 3.

Procedure. A total of one hundred thirty-six undergraduate students participated in this

study in exchange for partial course credit. The procedure followed that of experiment 3 except that all participants were presented with either the outcome or the process accountability manipulation instructions at the beginning of the study.

Results

Category 1: Attempt-to-align false recalls. Regression analysis revealed a significant two-way interaction between expertise and accountability ($\beta = .22, t = 2.09, p < .05$). To further examine the interaction we plotted the graphs at one standard deviation above and below the mean of the expertise scale (see figure 2). Individuals under process accountability generated a higher proportion of attempt-to-align false recalls at higher levels of expertise (M = 6.29%) than at lower levels (M = 1.35%; $\beta = 1.95, t = 3.00, p < .01$). However, no difference was observed between lower expertise (M = 2.75%) and higher expertise (M = 2.86%) under outcome accountability condition ($\beta = .05, t < 1$).

Contrast analysis indicated no difference in the proportion of attempt-to-align false recalls generated between the process (M = 1.35%) and outcome (M = 2.75%) accountability conditions when expertise was low (t < 1). However, in line with H4A, when expertise was high, more false recalls were generated under process accountability (M = 6.29%) than under outcome accountability condition (M = 2.86%; $\beta = 3.43$, t = 2.11, p < .05).

Category 2: Incorrect value recalls. As expected, for recalls unrelated to attempts to align attributes, no significant two-way interaction was observed ($\beta = .09, t < 1$).

Category 3: Confusions. There were no main or interactive effects for confusions (all *t*'s < 1).

Insert figure 2 about here

Total and correct recalls. No significant interaction between expertise and accountability appeared for either total recalls or correct recalls (all t's < 1). Consistent with previous studies, again we observed a significant main effect of expertise for both total recalls (β = .17, t = 3.63, p < .001) and correct recalls (β = .17, t = 3.33, p < .01).

Evaluation. Evaluation indices were created for option A ($\alpha = .79$) and option B ($\alpha = .83$). An overall evaluation index was generated exactly as in experiment 3.

A significant two-way interaction between expertise and accountability was observed $(\beta = ..33, t = .3.10, p < .01)$ (see figure 3). Contrast analysis conducted at 1 S.D. above and below the mean of the expertise scale indicated that under process accountability option B was evaluated higher when expertise was low (M = 2.19) than when it was high ($M = ..35; \beta = ..73, t = ..2.51, p < ..05$). The opposite pattern appeared for outcome accountability. Option B was evaluated marginally significantly higher when expertise was high (M = 2.42) as compared to when it was low ($M = 1.09; \beta = ..53, t = 1..87, p = ..06$). Further contrast analysis indicated that, as predicted by H4B, at higher levels of expertise outcome accountable respondents evaluated option B (M = 2.42) significantly higher than process accountable respondents ($M = ..35; \beta = ..2.07, t = ..2.86, p < ..01$). However, at lower levels of expertise there was no significant difference between evaluations by outcome accountable (M = 1.09) vs. process accountable respondents ($M = ..35; \beta = ..2.07, t = ..2.86, p < ..01$). However, at lower levels of expertise there was no significant difference between evaluations by outcome accountable (M = 1.09) vs. process accountable respondents ($M = ..2.19; \beta = 1..11, t = 1.53, ns$).

Insert figure 3 about here

Discussion

The results from study 4 provide insight into the manner in which feelings of accountability affect experts. The findings of the experiment suggest that the feelings of accountability that experts experience in memory-based product comparison lead them to focus on their decision process, rather than simply on the outcome. Although prior research has shown that process accountability can improve decision making (Lerner and Tetlock 1999), our study suggests that expertise may be a moderator of this general effect. We found that for experts a focus on process leads them to commit more attempt-to-align errors and consequently lower the quality of their decisions. Consistent with previous work showing that medical experts sometimes have difficulty articulating their decision processes (Wimmers et al. 2005), it appears that consumer experts may be better off simply trying to reach a correct outcome without concentrating on how it was reached.

For both experiments 3 and 4, one might question whether experts were truly less normative in their evaluations. Given their knowledge of the category, it is possible that experts' false recalls were in fact correct inferences. If so, their reduced evaluation of option B could in fact be normative. However, as described in experiment 3, our stimuli were carefully constructed so that option B was perceived to be superior to option A; thus, a reduced evaluation of option B would reflect some impairment in judgment. Moreover, the fact that evaluation of option B in experiment 4 varied as a function of the type of accountability puts this possibility further in doubt. If the false recalls were in fact correct, there would be no reason to expect a change in the evaluation of option B under process versus outcome accountability.

Although the evidence is supportive of a decrease in judgment quality for experts as a function of attempt-to-align false recalls, it is important to qualify this claim somewhat. In our experimental paradigm, participants evaluated both options individually and we then subtracted the score for option A from the score for option B. Although we argue that a lower score on this difference measure indicates a deviation from the normative baseline for experts, any positive value does indicate awareness that B is the superior option at least to some degree. Thus, we cannot unequivocally conclude that expertise would always result in a suboptimal choice. We can, however, attest to a lowered decision quality since evaluations of Option B decreased significantly with the higher attempt-to-align errors that are concomitant with greater expertise.

GENERAL DISCUSSION

The four experiments presented here contribute to research on expert memory in four ways. First, we demonstrated that in memory-based product comparisons, experts are prone to making attempt-to-align false recalls, errors that occur as a result of attempts to make features more comparable. Second, we provided evidence that these false recalls stem from a greater sense of accountability for the decision process that experts naturally feel, relative to novices. Accordingly, third, we show that relieving experts of process accountability reduces experts' errors. Finally, we showed that such recalls can lead experts to commit suboptimal product judgments. Although prior research has examined issues of expertise and accountability in consumer behavior (Kardes et al. 2004), our work is the first to demonstrate the effects of accountability on expert memory performance in a product comparison context.

Experiment 1 showed that when recalling features from products presented in a comparison context, experts made more spontaneous and attempt-to-align false recalls than novices. The experiment also tested and found support for the proposition that experts naturally feel a greater sense of accountability for their responses than novices, and demonstrated that this sense of accountability mediated attempt-to-align false recalls but not other types of false recalls.

Experiment 2 provided further support for the accountability explanation, showing that when experts were explicitly made to feel accountable for their responses, their false recalls were equivalent to when they were given no specific instructions. Importantly, we also showed that relieving experts of accountability reduced their tendency to commit attempt-to-align false recalls but not their tendency to commit other types of false recalls. This supports our proposition that attempt-to-align false recalls occur as a function of attempts to make product options comparable, rather than solely as a result of a highly developed schema. It is, we argue, the intersection of accountability and a well-developed schema that leads to susceptibility to attemptto-align false recalls.

Experiment 2 also demonstrated that involvement alone could not account for the type of false recalls observed. Prior research has shown that involvement can lead to a focus on nonalignable features (Zhang and Markman 2001) but has not examined memory errors and has not considered the roles of expertise or accountability. Our findings suggest that for experts, involvement and accountability operate differently. Not surprisingly, experts are more involved in tasks related to their domain of expertise, but this higher level of involvement does not appear

to lead directly to an increase in attempt-to-align false recalls. Rather, it is a feeling of being accountable that leads to these types of errors. This is consistent with prior research suggesting that involvement relates more to personal relevance whereas accountability is linked more to concerns about self presentation and judgment error (Lee et al. 1999).

In experiment 3 we examined how false recalls impacted subsequent product evaluation. We found that the relationship between expertise and product evaluation was mediated by attempt-to-align false recalls. That is, an increase in attempt-to-align false recalls, but not other types of false recalls, led to less normative product evaluations. It appears that the false recalls may have altered the influence of the nonalignable features in the judgment task (Houston et al. 1991) resulting in the dominant alternative appearing less dominant.

Our final experiment shed light on the process by which accountability contributes to experts' attempt-to-align false recalls. Specifically we determined that the accountability experts naturally feel relates to being able to explain their decision process, rather than to reach the correct outcome. When experts were made to feel accountable for the outcome alone and hence did not have to be concerned with explaining how they made their decision, their attempt-toalign false recalls were reduced and decision quality was improved.

We certainly do not claim that expertise will always be a hindrance in memory-based decision making. Of course, in many cases a well-developed schema will aid an expert in memory-based choices. However, in memory based product comparisons, false memories may bias consumers' product judgments because they can be confused with actual experiences, particularly when the false recalls are related or similar to the actual perceptions (Johnson 2006; Kardes 1988). We suspect that such biases will be most likely to occur in product categories that have numerous optional features (e.g., cars, financial products, cameras, etc.), categories in

which the technology may change rapidly (e.g., cellular phones, computers, etc.), or categories in which product options may have more nonalignable than alignable features (e.g., vacations, spa packages, etc.). In such cases, experts' memories may prevent them from conducting an accurate assessment of product options, resulting in biased and impaired judgments.

Accountability is not a factor in the false memory effects studied in previous research (e.g., Roediger and McDermott 1995). Thus, our work is the first to demonstrate the mediating role of accountability in the creation of false memories. We show that for experts, accountability leads to a focus on the decision process rather than a focus on the outcome, and this focus results in more attempt-to-align false recalls and ultimately poorer decision quality.

Across all experiments we operationalized expertise using subjective judgments rather than an objective measure, as we believed that subjective expertise is the more appropriate measure in our context. Our guiding premise is that consumer experts' feelings of accountability for the information they provide lead to false recalls. These feelings should stem from their own beliefs in their expertise, rather than how much they actually know. Nonetheless, it is useful to consider how an objective measure would affect our results. We expect that the pattern of effects would be quite similar, as the correlation between subjective and objective knowledge, while far from perfect (Brucks 1985), is positive. However, if self-proclaimed experts actually lacked sufficient objective knowledge, they would be unable to generate the attempt-to-align false recalls despite their best efforts and hence their recalls would mirror those of the novices.

Our results are, of course, constrained by the context in which we investigated our research question. We employed a free recall task as our measure of memory performance. However, it is entirely plausible that consumers would have cues available to them to aid their comparison process. An important question is whether the pattern of results would remain the same if participants were given a recognition task rather than a recall task. We speculate that indeed, the results would hold, and in some respects may get stronger. Given that the errors we studied occurred as a result of trying to make the options more comparable, it is likely that in a recognition task, expert consumers would record a high number of "false positives," recognizing features as having been presented previously for both options when in fact they were originally presented only for one. Future research could examine this important issue.

The experimental context also constrained the recall procedure. We provided participants with empty text boxes, side-by-side on a computer screen and asked them to recall as many features for each option as they could. Because we were interested in their free recall abilities, we did not control how they completed the recall task (i.e., feature by feature for both options, all the features for one option followed by all the features for the other, etc.). However, it is interesting to speculate whether altering the recall task by forcing participants to either recall the products one at a time or recall the features one at a time would alter the results. It is likely that a focus on the comparability of the options (i.e., recalling features for both options one feature at a time) would exacerbate our effects, whereas focusing on the individual products (i.e., recalling all the features for option A and then all the features for option B) would mitigate the attempt-to-align errors.

A final constraint of our experimental paradigm was the stimulus we chose. Video games are a predominantly male activity and as such, males tend to have greater domain expertise. Although we provided statistical evidence that our results were due to expertise and not the result of a gender effect, additional research should confirm that our effects generalize to products for which gender is not a concern.

The goals of the current work were to demonstrate that experts are more prone to

attempt-to-align false recalls, and that these false recalls occur as a result of feelings of accountability that experts naturally experience. However, an important related issue is the extent to which these false recalls are constructive or reconstructive. That is, are experts making inferences at the time of original processing which they later confuse with the actual stimulus, or are the errors occurring at the time of retrieval? Based on the nature of the errors we encountered it seems that some of the recall errors, like incorrect value and confusions, for which encoding inferences were not possible, are likely retrieval errors. However, for attempt-to-align false recalls, the errors could have occurred at encoding or at retrieval.

Although the question of whether attempt-to-align false recalls are a function of encoding or retrieval is outside the scope of the current research, it is interesting to consider it in light of the "Selective Construction and Preservation of Experiences" (SCAPE) model (Mantonakis, Whittlesea and Yoon 2008). This model proposes that all memory experiences involve the operation of reconstructive inference processes (see also Braun 1999; Braun et al. 2002). Pieces of information come to mind and must be evaluated as to whether they are significant for the current task or context. In our experimental paradigm, when participants are asked to recall product information, various features come to mind and must be evaluated for appropriateness. Because novices have few product representations in memory, any features that come to mind are likely evaluated as relevant and consequently are recalled. Experts, on the other hand, have numerous product representations already in memory; thus, they must decide if the features that come to mind are relevant. One measure of relevance could be whether the piece of information helps them in their comparative task. Because features that increase alignability aid in the task, features that were not originally presented but align with a presented feature (i.e., attempt-toalign false recalls) may be judged as significant and hence recalled. Although our research does

not speak directly to this issue, we believe this is an important question and could be a fruitful direction for future research.

We have highlighted one mechanism underlying attempt-to-align false recalls, but others may exist. In addition, it would be interesting to investigate whether there are conditions under which the other types of false recalls (i.e., incorrect value and confusions) would lead to impaired product judgment. Future research could also investigate ways to improve expert performance. Given their more developed knowledge there may be ways to alert experts to the risk of these attempt-to-align errors and learn techniques to correct for them.

APPENDIX A

OPTION FEATURES FOR THE VIDEOGAME STIMULI (EXPERIMENT 1 AND 2)*

Option A	Option B	Fcature Typ e
No hard drive	20 GB Hard Drive	
HDMI output option not available	HDMI Output available	Aligned
Dolby analog sound output	Analog sound output is not Dolby (is Stereo)	Features
Live and recorded TV support	Live and recorded TV support not available	
Controller doesn't have motion sensor	Free online service like ability to download full-	
functionality	length movies and TV shows	
720p video resolution	Bluetooth 2.0 wireless networking	Non-Aligned
512 MP Main System PAM shared with GPU	Does not support streaming audio or video	Features
512 MID Main System RAW shared with OF C	from PCs	
512 MB Video RAM	Controller doesn't have rumble functionality	

* In the stimuli presented to participants these features were randomized within option and Feature Type was not indicated.

APPENDIX B

FOUR VERSIONS OF STIMULI USED EXPERIMENT 1-2

Option A	Option B			
Version 1				
720p video resolution	20 GB Hard Drive			
512 MB Video RAM	Live and recorded TV support not available			
Dolby analog sound output	Analog sound output is not Dolby (is Stereo)			
HDMI output option not available	Bluetooth 2.0 wireless networking			
No hard drive	Free online service like ability to download full-length movies and TV shows			
512 MB Main System RAM shared with GPU	Controller doesn't have rumble functionality			
Live and recorded TV support	Does not support streaming audio or video from PCs			
Controller doesn't have motion sensor functionality	HDMI Output available			
Vers	lon 2			
Wireless networking option not available	Controller has motion sensor functionality			
Supports streaming audio and video from PCs	256 MB Video RAM			
512 MB Main System RAM shared with GPU	256 MB Main System RAM not shared with GPU			
720n video resolution	Near Complete compatibility with previous generation			
	games			
Controller doesn't have motion sensor functionality	Shall support next generation DVD formats			
Controller has rumble functionality	1080p video resolution			
512 MB Video RAM	Cannot upscale DVD movies to HD resolutions			
Online service like ability to download full-length	Cannot improve legacy (previous generation) game			
movies and TV shows available for a yearly fee	experience by improvement of resolutions or graphics			
Vers	ion 3			
	Free online service like ability to download full-length			
Partially compatible with previous generation games	movies and TV shows			
Upscales DVD movies to HD resolutions	Controller doesn't have rumble functionality			
Online service like ability to download full-length movies and TV shows available for a yearly fee	Does not support streaming audio or video from PCs			
Wireless networking option not available	HDMI Output available			
Supports streaming audio and video from PCs	20 GB Hard Drive			
Improves legacy (previous generation) game experience through improved resolutions or graphics	Live and recorded TV support not available			
Controller has rumble functionality	Analog sound output is not Dolby (is Stereo)			
Does not support next generation DVD formats	Bluetooth 2.0 wireless networking			
Version 4				
HDMI output option not available	Shall support next generation DVD formats			
Partially compatible with previous generation games	Cannot improve legacy (previous generation) game experience by improvement of resolutions or graphics			
Upscales DVD movies to HD resolutions	Cannot upscale DVD movies to HD resolutions			
Dolby analog sound output	1080p video resolution			
Does not support next generation DVD formats	Controller has motion sensor functionality			
Live and recorded TV support	256 MB Video RAM			
No hard drive	256 MB Main System RAM not shared with GPU			
Improves legacy (previous generation) game	Near Complete compatibility with previous generation			
experience through improved resolutions or graphics	games			

APPENDIX C

RECALL RATES AT -1/+1 STANDARD DEVIATION FROM MEAN EXPERTISE

Study	Variable	-1 SD		+1 SD			
		Effect of expertise		Effect of expertise			
	Total Recall	7.32			9.23		
	Correct Recall		6.66		8.06		
	Correct Aligned		4.07		4.44		
1	Correct Nonaligned		2.59		3.63		
1	Total Incorrect Recall		0.66		1.16		
	Attempt to Align		0.13		0.44		
	Percent Attempt to Align		1.54%		5.30%		
	Incorrect Value	0.29			0.58		
	Percent Incorrect Value		4.09%		7.51%		
	Confusions		0.17		0.14		
		Control	Acc	Unacc	Control	Acc	Unacc
	Total Recall	6.03	5.79	6.48	9.31	8.11	8.31
	Correct Recall	5.68	5.19	5.51	7.65	6.24	7.53
	Correct Aligned	3.43	2.89	3.14	4.62	3.75	4.37
	Correct Nonaligned	2.25	2.29	2.37	3.03	2.49	3.16
2	Total Incorrect Recall	0.35	0.60	0.97	1.66	1.87	0.79
	Attempt to Align	0.12	0.11	0.32	0.67	0.99	0.14
	Percent Attempt to Align	1.65%	1.20%	5.60%	7.12%	10.52%	1.89%
	Incorrect Value	0.25	0.26	0.33	0.40	0.56	0.46
	Percent Incorrect Value	6.35%	3.65%	5.22%	4.48%	6.27%	5.61%
	Confusions	-0.06	0.14	0.33	0.65	0.27	0.26
	Total Recall	9.13		12.79			
	Correct Recall		8.71		10.82		
	Correct Aligned		3.72		4.60		
	Correct Nonaligned		1.77		2.51		
	Recall Same		3.22		3.72		
3	Total Incorrect Recall	0.43			1.97		
	Attempt to Align	0.08			1.17		
	Percent Attempt to Align	2.57%			9.35%		
	Incorrect Value	0.23			0.64		
	Percent Incorrect Value	1.89%			5.94%		
	Confusions	0.23			0.21		
		Outcome	Acc P	rocess Acc	Outcome	Acc P	rocess Acc
	Total Recall	7.85		8.35	11.40 11		11.43
	Correct Recall	7.24		7.94	10.52 10.00		10.00
	Correct Aligned	3.03		3.08	4.33 3.32		3.32
4	Correct Nonaligned	1.61		1.39	2.62 2.53		2.53
	Recall Same	2.60		3.47	3.57 4.15		4.15
	Total Incorrect Recall	0.61		0.41	0.88 1.43		1.43
	Attempt to Align	0.21		0.12	0.30 0.73		0.73
	Percent Attempt to Align	2.75%)	1.35%	2.86% 6.29%		6.29%
	Incorrect Value	0.28		0.16	0.45		0.60
	Percent Incorrect Value	4.27%)	1.42%	6.06%)	6.20%
Confusions 0.13			0.13	0.05		0.10	

APPENDIX D:

OPTION FEATURES OF THE VIDEOGAME STIMULI USED IN EXPERIMENT 3 AND 4*

		Feature
Option A	Option B	Туре
20 GB Hard Drive	60 GB Hard Drive	
Live and recorded TV support not available	Live and recorded TV support	Aligned
256 MB Video RAM	512 MB Video RAM	Features
Analog sound output is not Dolby (is Stereo)	Dolby analog sound output	
3.2GHz Processor	HDMI output option not available	
1080p native video resolution	Controller doesn't have motion sensor	
	functionality	Non-Aligned
Upscales DVD movies to HD resolutions	8.0 GBps Memory Bandwidth	Features
	Partially compatible with previous generation	
Bluetooth 2.0 wireless networking	games	
Supports 4 wireless controllers	Supports 4 wireless controllers	
Free online service like ability to download full-	Free online service like ability to download full-	Common
length movies and TV shows	length movies and TV shows	Eestures
Controller has rumble functionality	Controller has rumble functionality	Teatures
12X dual-layer DVD Optical drive	12X dual-layer DVD Optical drive	

* In the stimuli presented to participants these features were randomized within option and Feature Type was not indicated.

REFERENCES

Aiken, Leona S. and Stephen G.West (1991), Multiple Regression, Newbury Park, CA: Sage.

- Alba, Joseph W. and Wesley J. Hutchinson (1987), "Dimensions of Consumer Expertise," *Journal of Consumer Research*, 13 (March), 411-54.
- Anderson, Craig A. and Brad J. Bushman (2001), "Effects of Violent Video Games on Aggressive Behavior, Aggressive Cognition, Aggressive Affect, Physiological Arousal, and Prosocial Behavior: A Meta-Analytic Review of the Scientific Literature," *Psychological Science*, 12 (September), 353-359.
- Baird, Raymond R. (2003), "Experts Sometimes Show More False Recall Than Novices: A Cost of Knowing Too Much," *Learning and Individual Differences*, 3, 349–55.
- Ballou, Brian (2001), "The Relationship between Auditor Characteristics and the Nature of Review Notes for Analytical Procedure Working Papers," *Behavioral Research in Accounting*, 13, 25-48.
- Bettman, James R., Eric Johnson, and John W. Payne (1991), "Consumer Decision Making," in *Handbook of Consumer Behaviour*, ed. T.S. Robertson and H. H. Kassarjian, Englewood Cliffs, NJ: Prentice Hall, 50–84.
- Biehal, Gabriel and Dipankar Chakravarti (1986), "Consumers' Use of Memory and External Information in Choice: Marco and Micro Perspectives," *Journal of Consumer Research*, 12 (March), 382-405.
- Black, Nick, Susan van Rooyen, Fiona Godlee, Richard Smith, and Stephen Evans (1998),
 "What Makes a Good Reviewer and a Good Review for a General Medical Journal?" *The Journal of the American Medical* Association, 280 (July), 231-233.

Braun, Kathryn A. (1999), "Postexperience Advertising Effects on Consumer Memory," Journal

of Consumer Research, 25 (March), 319–334.

- Braun, Kathryn A., Rhiannon Ellis, and Elizabeth F. Loftus (2002), "Make My Memory: How Advertising Can Change Our Memories of the Past," *Psychology and Marketing*, 19 (January), 1-23.
- Brucks, Merrie (1985), "The Effects of Product Class Knowledge on Information Search Behavior," *Journal of Consumer Research*, 12 (June), 1-16.
- Burke, Raymond R. and Thomas K. Srull (1988), "Competitive Interference and Consumer Memory for Advertising," *Journal of Consumer Research*, 15, No.1 (June), 55-68.
- Carnagey, Nicholas L. and Craig A. Anderson (2005), "The Effects of Reward and Punishment in Violent Video Games on Aggressive Affect, Cognition, and Behavior," *Psychological Science*, 16 (11), 882-889.
- Castel, Alan D., David P. McCabe, Henry L. Roediger, and Jeffrey L. Heitman (2007), "The Dark Side of Expertise: Domain-Specific Memory Errors," *Psychological Science*, 18 (January), 3-5.
- Chase, William G. and Herbert A. Simon (1973), "Perception in Chess," *Cognitive Psychology*, 4 (1), 55-81.
- Coley, John, Shafto, Patrick, Stepanova, Olga, and Elizabeth Baraff (2005), "Knowledge and Category-Based Induction," in *Categorization Inside and Outside the Laboratory: Essays in Honor of Douglas L. Medin*, eds. Woo-kyoung Ahn, Robert L. Gladstone, Bradley C. Love, Arthur B. Markman, and Phillip Wolff, APA: Decade of Behavior Series.
- Cowley, Elizabeth and Andrew A. Mitchell (2003), "The Moderating Effect of Product Knowledge on the Learning and Organization of Product Information", *Journal of Consumer Research*, 30 (December), 443-454.

- Cowley, Elizabeth and Mitchell Andrew A (2005), "What Do Novice Consumers Remember?" Advances in Consumer Research, 32, 437-38.
- Evans, Arthur T., McNutt, Robert A., Fletcher, Suzanne W., and Robert H. Fletcher (1993), "The Characteristics of Peer Reviewers Who Produce Good-Quality Reviews," *Journal of General Internal Medicine*, 8 (8), August, 422-428.
- Feng, Jing, Ian Spence, and Jay Pratt (2007), "Playing an Action Video Game Reduces Gender Differences in Spatial Cognition." *Psychological Science*, 18 (10), 850-855.
- Houston, David A., Steven J. Sherman, and Sara M. Baker (1991), "Feature Matching, Unique Features, and the Dynamics of the Choice Process: Predecision Conflict and Postdecision Satisfaction," *Journal of Experimental Social Psychology*, 27, 411-430.
- Huffman, Cynthia and Michael J. Houston (1993), "Goal-Oriented Experiences and the Development of Knowledge," *Journal of Consumer Research*, 20 (September), 190-207.
- Johnson, Marcia K. (2006), "Memory and Reality," *American Psychologist*, 61 (November), 760-771.
- Kardes, Frank (1988), "A Nonreactive Measure of Inferential Beliefs," *Psychology & Marketing*, 5 (Fall), 273-286.
- Kardes, Frank R., A. V. Muthukrishnan, and Vladimir Pashkevich (2004), "On the Conditions Under Which Experience and Motivation Accentuate Bias in Intuitive Judgment," in *The Routines of Decision Making*, eds. Tilmann Betsch and Susanne Haberstroh, Mahwah, NJ: Lawrence Erlbaum Associates, 139-156.
- King, Charles W. and John O. Summers (1970), "Overlap of Opinion Leadership across Consumer Product Categories," *Journal of Marketing Research*, 7 (February), 43-50.

- Kray, Laura and Richard Gonzalez (1999), "Differential Weighting in Choice versus Advice: I'll do this, you do that," *Journal of Behavioral Decision Making*, 12, 207-217.
- Langer, Ellen J. and Lois G. Imber (1979), "When Practice Makes Imperfect: Debilitating Effects of Overlearning," *Journal of Personality and Social Psychology*, 37 (November), 2014-25.
- Lee, Hanjoon, Paul M. Herr, Frank R. Kardes, and Chankon Kim (1999), "Motivated Search: Effects of Choice Accountability, Issue Involvement, and Prior Knowledge on Information Acquisition and Use," *Journal of Business Research*, 45 (May), 75-88.
- Leonard-Barton, Dorothy (1985), "Experts as Negative Opinion Leaders in the Diffusion of a Technological Innovation" *Journal of Consumer Research*, 11 (March), 914-26.
- Lerner, Jennifer S. and Philip E. Tetlock (1999), "Accounting for the Effects of Accountability," *Psychological Bulletin*, 125 (March), 255-75.
- Liberman, Nira and Jens Forster (2006), "Inferences from Decision Difficulty," *Journal of Experimental Social Psychology*, 42, 290-301.
- Lynch, John G., Jr. and Thomas K. Srull (1982), "Memory and Attentional Factors in Consumer Choice: Concepts and Research Methods," *Journal of Consumer Research*, 9 (June), 18-37.
- Mackiewicz, Jo (2010), "Assertions of Expertise in Online Product Reviews," *Journal of Business and Technical Communication*, 24 (1), 3-28.
- MacKinnon, David P., Chondra M. Lockwood, and Jason Williams (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39 (1), 99–128.

Mantonakis, Antonia, Bruce W. A. Whittlesea, and Carolyn Yoon (2008), "Consumer Memory,

Fluency, and Familiarity," in Handbook of Consumer Psychology, ed. Curtis P.

Haugtvedt, Paul M. Herr, and Frank R. Kardes, New York, NY: Psychology Press, 77-102.

- Marchant, Garry, John Robinson, Urton Anderson, and Michael Schadewald (1991), "Analogical Transfer and Expertise in Legal Reasoning," *Organizational Behavior and Human Decision Processes*, 48 (April), 272-90.
- Markman, Arthur B. and Jeffrey Loewenstein (2010), "Structural Comparison and Consumer Choice," *Journal of Consumer Psychology*, 20 (April), 126-137.
- Markman, Arthur B. and Douglas L. Medin (1995), "Similarity and Alignment in Choice," Organizational Behavior and Human Decision Processes, 63 (August), 117–130.
- Mather, Mara, Marisa Knight, and Michael McCaffrey (2005), "The Allure of the Alignable:
 Younger and Older Adults' False Memories of Choice Features," *Journal of Experimental Psychology: General*, 134 (February), 38-51.
- Park, C. Whan, David L. Mothersbaugh, and Lawrence Feick (1994), "Consumer Knowledge Assessment," *Journal of Consumer Research*, 21 (June), 71–82.
- Patel, Vimla L. and Guy J. Groen (1991), "The General and Specific Nature of Medical
 Expertise: a critical look," in *Toward a General Theory of Expertise, Prospects and Limits*, ed. KA Ericsson, J Smith, Cambridge: Cambridge University Press, 93-125.
- Preacher, Kristopher J. and Andrew F. Hayes (2008), "Asymptotic and Resampling Strategies for Assessing and Comparing Indirect Effects in Multiple Mediator Models," *Behavior Research Methods*, 40 (3), 879-91.
- Roediger, Henry L. III and Kathleen B. McDermott (1995), "Creating False Memories: Remembering Words not Presented in Lists," *Journal of Experimental Psychology*:

Learning, Memory, and Cognition, 21 (July), 803-814

- Roehm, Michelle L. and Brian Sternthal (2001), "The Moderating Effect of Knowledge and Resources on the Persuasive Impact of Analogies," *Journal of Consumer Research*, 28 (September), 257-72.
- Sanbonmatsu, David M., Frank R. Kardes, and Bryan D. Gibson (1991), "The Role of Attribute Knowledge and Overall Evaluations in Comparative Judgment," *Organizational Behavior and Human Decision Processes*, 48 (February), 131-146.
- Schmidt, Henk G. and Henny P. A. Boshuizen (1993), "On Acquiring Expertise in Medicine," *Educational Psychology Review*, 5 (3), 205–21.
- Siegel-Jacobs, Karen and J. Frank Yates (1996), "Effects of Procedural and Outcome Accountability on Judgment Quality," *Organizational Behavior and Human Decision Processes*, 65 (January), 1-17.
- Simonson, Itamar and Peter Nye (1992), "The Effect of Accountability on Susceptibility to Decision Errors," Organizational Behavior and Human Decision Processes, 51 (April), 416-46.
- Simonson, Itamar and Barry M. Staw (1992), "Deescalation Strategies: A Comparison of Techniques for Reducing Commitment to Losing Courses of Action," *Journal of Applied Psychology*, 77 (August), 419-26.
- Simonson, Itamar and Amos Tversky (1992), "Choice In Context: Tradeoff Contrast And Extremeness Aversion," *Journal of Marketing Research*, 29 (August), 281–295.
- Sjoberg, L. (1972), "A Cognitive Theory of Similarity," *Goteborg Psychological Reports*, 2 (10), 1-23.

Sternberg, Robert J. (1997), "Cognitive Conceptions of Expertise," in Expertise in Context:

Human and Machine, ed. P. J. Feldovich, K. M. Ford and R. R. Hoffman, Menlo Park, California: AAAI Press/The MIT Press.

- Tetlock, Philip E. (1983), "Accountability and Complexity of Thought," *Journal of Personality and Social Psychology*, 45 (July), 74-83.
- Tetlock, Philip E. (1986). "A Value Pluralism Model of Ideological Reasoning," *Journal of Personality and Social Psychology*, 50 (4), 819-827.
- Tetlock, Philip E. and Richard Boettger (1989), "Accountability: A Social Magnifier of the Dilution Effect," *Journal of Personality and Social Psychology*, 57 (3), 388-98.
- Underwood, Benton J. (1965), "False Recognition Produced by Implicit Verbal Responses," Journal of Experimental Psychology, 70 (July), 122-29.
- Voss, James F., Gregg T. Vesonder, and George J. Spilich (1980), "Text Generation and Recall by High-Knowledge and Low-Knowledge Individuals," *Journal of Verbal Learning and Verbal Behaviour*, 19 (December), 651-67.
- Wang, Jing and Robert S. Wyer, Jr. (2002), "Comparative Judgment Processes: The Effects of Task Objectives and Time Delay on Product Evaluations," *Journal of Consumer Psychology*, 12 (4), 327-340.
- Wimmers, Paul F, Henk G. Schmidt, Peter P. J. L. Verkoeijen, and Margje W. J.van de Wiel (2005), "Inducing Expertise Effects in Clinical Case Recall," *Medical Education*, 39 (September), 949-57.
- Zhang, Shi and Arthur B. Markman (1998), "Overcoming the Early Entrant Advantage: The Role of Alignable and Nonalignable Differences," *Journal of Marketing Research*, 35 (November), 413-426.

Zhang, Shi and Arthur B. Markman (2001), "Processing Product Unique Features: Alignability

and Involvement in Preference Construction," *Journal of Consumer Psychology*, 11 (1), 13-27.

TABLE 1ERROR CLASSIFICATION

Description of Recall Error	Error Classified as	Example		
Errors Related to Aligned Features	•			
One of the two feature values recalled is incorrect (i.e., not presented originally for either option).	Incorrect Value	Originally presented : 60 GB hard drive for option A, 40 GB hard drive for option B Recalled: 60 GB hard drive for option A, 20 GB hard drive for option B		
Both feature values recalled are incorrect (i.e., not presented originally for either option).	Incorrect Value	Originally presented: 60 GB hard drive for option A, 40 GB hard drive for option B Recalled : 80 GB hard drive for option A, 20 GB hard drive for option B		
Recalled feature values are correct, but switched between the options.	Confusions	Originally presented: 60 GB hard drive for option A, 40 GB hard drive for option B Recalled: 40 GB hard drive for option A, 60 GB hard drive for option B		
Option A's feature value is recalled for Option B or vice versa (only 1 feature is recalled)	Confusions	Originally presented: 60 GB hard drive for option A, 40 GB hard drive for option B Recalled: 40 GB hard drive for option A, nothing recalled for option B		
Errors Related to Non-Aligned Features	-			
Incorrect value of presented feature recalled for the correct option and no recall with respect to this feature for the other option.	Incorrect Value	Originally presented: 512 MB RAM for option A; nothing mentioned for option B Recalled: 256 MB RAM for option A, nothing for option B		
Incorrect value of the presented feature recalled for the incorrect option. No recall with respect to this feature for the correct option.	Incorrect Value	Originally presented : 512 MB RAM for option A; nothing mentioned for option B Recalled: nothing for option A; 256 MB RAM for option B		
Correct value of the presented feature recalled, but for the incorrect option. No recall with respect to this feature for the correct option.	Confusions	Originally presented: 512 MB RAM for option A; nothing mentioned for option B Recalled: nothing for option A; 512 MB RAM for option B		
Correct value of the presented feature recalled for the correct option and a matching feature is recalled for the other option.	Attempt-to- align	Originally presented: 512 MB RAM for option A; nothing mentioned for option B Recalled: 512 MB RAM for option A; 256 MB RAM for option B		
Incorrect value of the presented feature recalled for the correct option and a matching feature is recalled for the other option.	Attempt-to- align	Originally presented: 512 MB RAM for option A; nothing mentioned for option B Recalled: 256 MB RAM for option A; 1024 MB RAM for option B		
Recall of features not originally presented for any of the options				
Recall a new feature for one of the options	Incorrect Value	Originally presented: N/A Recalled: online multi user capability for option A; nothing for option B		
Recall matching features for both the options	Attempt-to- align	Originally presented : N/A Recalled: online multi user capability for option A; online multi user capability for option B		
Random comments provided that were unrelated to videogame consoles, i.e. no mention of technical features	Random	Originally presented : N/A Recalled: I don't like video games.		

FIGURE 1:

PROPORTION OF ATTEMPT-TO-ALIGN FALSE RECALLS GENERATED UNDER THREE ACCOUNTABILITY CONDITIONS AT \pm 1 S.D. OF THE MEAN (EXPERIMENT 2)



FIGURE 2: PROPORTION OF ATTEMPT-TO-ALIGN FALSE RECALLS GENERATED FOR OUTCOME VS. PROCESS ACCOUNTABILITY AT ± 1 S.D. OF THE MEAN (EXPERIMENT 4)



FIGURE 3: OPTION EVALUATION UNDER OUTCOME VS. PROCESS ACCOUNTABILITY AT ± 1 S.D. OF THE MEAN EXPERTISE (EXPERIMENT 4)*



* A higher value means better evaluation for option B while a rating of zero means no difference between the evaluation of option A and option B