Across four studies, including one involving an actual monetary decision, the authors demonstrate that forming a product consideration set by excluding versus including alternatives induces consumers to place more weight on ethical attributes, such as company labor practices and animal testing. This nonnormative difference reflects a compatibility between exclusion and ethics, and it holds regardless of attribute framing or consumer emotion. The authors also find that consumers judge others' behavior more negatively for excluding ethical products than for including ethical products. These results have implications for the marketing of ethical products, both specifically (e.g., it is important to encourage exclusion modes) and generally (e.g., the failure to consider ethical products may reflect seemingly minor contextual issues guiding the decision process and not consumer disinterest in ethical issues).

Keywords: ethics, response mode compatibility, consideration sets, include/exclude, decision making

Ethical Decisions and Response Mode Compatibility: Weighting of Ethical Attributes in Consideration Sets Formed by Excluding Versus Including Product Alternatives

One of the most pervasive and puzzling inconsistencies in human behavior is the discrepancy between stated values and actual behavior. Psychologists have studied this attitude–behavior link across many contexts (e.g., Eagly and Chaiken 1993) and have paid special attention to instances in which people appear to be practicing “moral hypocrisy” (Batson et al. 1999) by not reflecting their supposed moral beliefs in their observed actions. For marketers, this puzzle takes a concrete form; as others have noted (e.g., Ehrich and Irwin 2005), it is surprising that products embodying commonly held values do not perform better in the marketplace. For example, why does furniture made from rainforest (versus tree farm) wood continue to sell? Why have market forces not eliminated animal testing in the cosmetics industry? Why are corporations able to continue to mistreat workers, often with no obvious market reaction? Ehrich and Irwin (2005) show that part of the problem is that consumers will not ask for ethical product information, even though they would use it if it were available.

For some product categories, however, ethical attribute information is readily available. In such cases, why might there be a discrepancy between values and behavior? Along with myriad other possibilities, such as hypocrisy (Batson et al. 1999), there may be contextual elements of the decision that guide consumers toward (or away from) considering the ethical possibilities.

Imagine being in the market for an electronic product, such as a cell phone. You survey the vast array of choices and, in doing so, discover that some of the phones are produced using child labor. Because the product category is large, you decide to establish a subset of phones you would
further consider. In forming this set, would you pay special attention to child labor, or would you screen on more attractive attributes, such as the price, performance, and style of the phone?

Now, instead of forming a set through inclusion, suppose that you formed a set only of the phones you would not further consider. The exclusion context may encourage you to explicitly identify the less ethical cell phones (the ones that are made using child labor) as phones to exclude from further consideration. Winnowing down a choice set in this way, by excluding unattractive options, may make it more difficult for consumers to ignore ethical attributes, even for those who might screen (in inclusion) solely on attributes such as price and performance. Perhaps because of the difficulty in prescreening on the basis of ethical attributes in an inclusion mode, the market often shields consumers from needing to make trade-offs between ethical and other attributes. Some examples of such screens are separate prescreened sections in larger retail stores (e.g., the organic or “natural” grocery section), entire retail locations dedicated to an ethical issue (e.g., fair-trade stores), and consumer-guide lists of companies and brands that reflect a particular ethical value (e.g., People for the Ethical Treatment of Animals [PETA’s] “cruelty-free” shopping guides).

In this set of studies, we establish that “exclusion” versus “inclusion” (Heller, Levin, and Goransson 2002; Levin, Huneke, and Jasper 2000; Levin et al. 2001) results in greater weighting of ethical attributes in consideration set formation, even though, normatively, the specific task used should not have any systematic influence. We also show that this effect is not due to alternative explanations, such as the positive/negative framing that influences choice or rejection among two alternatives (Shafir 1993) or the added negative emotion in exclusion (Luce 1998; Luce, Payne, and Bettman 2000). Rather, we argue that the inclusion and exclusion tasks are differentially compatible (Fischer et al. 1999; Fischer and Hawkins 1993; Tversky, Sattath, and Slovic 1988) with ethical attributes. These results have implications for our understanding of how ethical attributes interact with decision modes and also add to the taxonomy of attributes and response mode compatibility. Most important, the results suggest that consumer consideration of ethical products is driven not only by motivational issues, such as hypocrisy and guilt, but also by simple cognitive issues, such as how context guides the decision in one direction versus another.

CONSIDERATION SETS

Consumer decisions involving several alternatives are often conceptualized as a two-stage process consisting of (1) prechoice screening and (2) choice (Beach and Mitchell 1990; Bettman and Park 1980; Nedungadi 1990; Ordóñez, Benson, and Beach 1999). Screening narrows the pool of options to a subset of all the available brands or alternatives in the population, commonly termed the “consideration set.” In the choice stage, consumers select an alternative or brand from the consideration set. Although marketers are interested in understanding the second stage, they must also understand how consideration sets are formed, for two reasons. First, being part of the consideration set is a necessary condition for being chosen (Desai and Hoyer 2000; Diehl 2005; Diehl, Kornish, and Lynch 2003; Nedungadi 1990). Second, the strategies used in the two stages can differ, resulting in differential weighting of variables in one versus the other stage (e.g., Chakravarti, Janiszewski, and Ulkümen 2006; Erdem and Swait 2004; Van Zee, Paluchowski, and Beach 1992). Given this prior research, consideration set formation must be treated as a distinct phase from ultimate choice, which is how we address it in this research. Several variables, such as usage situation (e.g., Desai and Hoyer 2000; Hutchinson, Ramam, and Mantrala 1994), assortment size (Heller, Levin, and Goransson 2002; Levin et al. 2001; Ordóñez, Benson, and Beach 1999; Payne, Bettman, and Johnson 1993), pioneering advantage (Kardes et al. 1993), and advertising (Mitra 1995; Mitra and Lynch 1995), have been shown to affect consideration set formation.

ETHICAL ATTRIBUTES IN CONSUMER DECISION MAKING

An ethical attribute is an attribute that reflects a person’s conscience (Baron and Spranca 1997; Ehrich and Irwin 2005; Irwin and Baron 2001). Products can have worse (e.g., they are made with child labor) or better (e.g., they are made without child labor) levels on an ethical attribute. Often, ethical attributes are linked to “protected” (Baron and Spranca 1997; Irwin and Baron 2001) or “sacred” (Tetlock, Kristel, and Elson 2000) values, which are values that people state they are unwilling (or at least reluctant) to trade off, no matter what the benefits of doing so may be. Being forced to trade off between attributes that reflect protected values and other attributes often generates substantial negative affect (Luce 1998).

Why might attributes that reflect a person’s ethical values (versus attributes without direct ethical implications, such as price) show the effects of such a simple variable as include/exclude? Although it might be expected that such values are immovable, research has shown that they can be especially vulnerable to contextual influence. Consumer actions driven by protected values, such as boycotting a firm that the consumer believes behaved egregiously, have been shown to be susceptible to influences such as message frame and social influence (e.g., Klein, Smith, and John 2004; Sen, Gürhan-Canli, and Morwitz 2001). In other purchase situations, such as evaluating items that vary in their environmental friendliness, the weighting of ethical attributes also seems to be especially prone to contextual effects. Consumers weight these attributes more in selling than in buying tasks (Irwin 1994), in choice than in pricing modes (Irwin and Baron 2001; Irwin et al. 1993), in separate than in joint evaluations (Irwin et al. 1993; Sunstein et al. 2002; Tenbrunsel et al. 2000), and when the attributes are presented singly than when they are presented jointly in a conjoint task (Irwin and Spira 1997).

In the case of inclusion versus exclusion, we expect that the context has a significant effect on the weighting of ethical attributes because of the differential compatibility of the attributes with the two modes. Compatibility has a long history in the study of response mode effects and has been shown to influence the weighting of ethical attributes (e.g., Irwin and Baron 2001; Irwin et al. 1993). In the next section, we review this history as it relates to our expectations and findings.
INCLUSION VERSUS EXCLUSION AND COMPATIBILITY WITH THE WEIGHTING OF ETHICAL ATTRIBUTES

Include/Exclude and Pick/Reject

Consideration sets are typically larger in exclusion modes than in inclusion modes (Heller, Levin, and Goranson 2002; Levin, Huneke, and Jasper 2000; Yaniv et al. 2002) because participants screen out more options when including than when excluding (Levin, Jasper, and Forbes 1998; Levin et al. 2001; Yaniv and Schol 2000). However, this effect does not necessarily presuppose differential weighting of attributes in the two modes. In binary choice, it has been established that asking participants to “pick” or “reject” one item can lead to different preference orders (Ganzach 1995; Meloy and Russo 2004; Shafir 1993; Wedell 1997) if some attributes are framed positively and some negatively. As Shafir (1993) explains, alternatives with both strongly positive and strongly negative aspects are chosen and rejected more often (than items with more consistently neutral attribute values) in binary choice/rejection tasks. The framing of the attributes differentially influences preference in the two modes because choice favors positive attributes and rejection favors negative attributes.

Although these framing effects imply that inclusion versus exclusion modes may guide preferences differently, they do not directly address the question of greatest interest to us: Regardless of framing, does consumers’ weighting of particular attributes differ by response mode? Particular attributes can be framed both positively and negatively (e.g., % fat versus % lean; Levin and Gaeth 1988); apart from frame, does the particular attribute matter?

Response Mode Compatibility

There is a large body of literature on how response modes can determine attribute weighting. For example, prominent attributes are weighted more in choice tasks than in other task modes, such as pricing and matching (Fischer et al. 1999; Fischer and Hawkins 1993; Irwin and Baron 2001; Lichtenstein and Slovic 1971; Tversky, Sattath, and Slovic 1988). This and other response mode effects have been explained using the notion of compatibility (Tversky, Sattath, and Slovic 1988): Response modes are differentially compatible with decision strategies.

Task Goal Compatibility

The type of compatibility most relevant to our hypotheses is “task goal compatibility,” or the choice of attribute weights that enable consumers to accomplish the goal of the particular task (Fischer et al. 1999; Fischer and Hawkins 1993). In differentiation tasks, such as choice, it is compatible with a decision maker’s goal simply to differentiate on the most important (prominent) attribute, given that the choices vary on this attribute; therefore, on average, this attribute is weighted more than others. Similarly, Hsee and colleagues (Hsee 1996; Hsee et al. 1999) show that the goal of differentiating between two items (e.g., in binary choice) results in more weight being given on evaluable, comparable attributes than does the goal of judging one item at a time.

We predict that ethical attributes will be weighted more in exclusion tasks because the goal of the exclusion mode is to indicate which alternatives a person does not want to consider further. Although, normatively, both modes make a statement about a person’s values, we argue that the expression of moral values feels more natural in exclusion than in inclusion: Ethical issues are compatible with the goal of indicating what a person does not want. Explicitly rejecting the “bad” items (the items unattractive on the ethical attribute) provides an indication of a person’s adherence to the ethical principle. This explicit rejection seems to match the reaction people have to values that are protected (or at least strongly held); the typical response to a particularly egregious ethical violation is to reject everything about it (Irwin 1994; Lichtenstein, Gregory, and Irwin 2007).

In inclusion, however, the goal is to indicate which options consumers most want to consider. Consumers may simply choose which alternatives to consider on the basis of the other attributes and, by default, not end up with the most ethical items in their consideration set. There is evidence for this difference in focus. Using verbal protocols, Irwin (1994) shows that selling modes encourage participants to consider how things might turn out badly in a transaction, while buying encourages participants to consider how things might turn out well. Thus, in selling modes, people favored the moral alternative more than in buying modes. Although a truly consistent ethical decision process would both reject negative instances from consideration and pick positive instances for definite inclusion, goal compatibility (Fischer et al. 1999; Fischer and Hawkins 1993) leads to different processes in the different modes.

Note that moral issues often take a form similar to exclusion (i.e., to reject that which is bad). For example, consider admonitions such as “Thou shalt not kill” (Ten Commandments); “First, do no harm” (a paraphrasing of the Hippocratic Oath); “Friends Don’t Let Friends Drive Drunk” (U.S. Department of Transportation campaign); and “Just Say No” (the 1980s and 1990s antidrug campaign). These and many other moral rules ask the decision maker to reject unacceptable alternatives and, by extension, not to reject acceptable alternatives. The act of rejection is key to the activation of the principle. It is likely that morality often takes this form because moral actions provide the avoidance of negative outcomes for the decision maker (e.g., being perceived as immoral).

In our conceptualization, ethicality is not equivalent to “importance”; ethical attributes are important, but they are more than that. As we show, product performance is important to the participants in our studies and to most consumers, but the importance of an attribute alone does not drive the proposed differential weighting of ethical attributes across inclusion and exclusion modes. Thus, we do not expect (and do not find) include/exclude effects for attributes that are important but not ethical, such as performance.

RESEARCH PLAN

In Studies 1 and 2, we show that ethical attributes are weighted more in consideration sets formed using exclusion versus inclusion. We also test another proposition based on our compatibility argument—namely, that there should be no difference in weighting for attributes without direct ethical implications. Thus, we do not expect to find differences in the weighting of nonethical attributes, such as price and performance, across response mode.
Although we theorize that compatibility is driving our expected result, there are two alternative explanations for this finding. First, attribute framing could induce our results, making them redundant with prior research (e.g., Shafir 1993). Ethical attributes often are expressed negatively (e.g., the number of exploited children, the amount of wasted environmental resources). To establish that the negativity of these attributes, not their frame, drives our results, we manipulate attribute frame in Study 2.

Second, differential weighting across the two modes might also be influenced by increased emotion in the exclusion mode that leads to avoidance of the ethical attribute. Loss modes, which may include exclusion contexts, induce negative emotion and concomitant coping strategies, such as avoidance (Luce 1998; Luce, Payne, and Bettman 2000). Thus, the increased emotion of exclusion and/or adding an ethical attribute to the decision process might cause participants to avoid making trade-offs and not to act on the ethically attractive options, resulting in these options being left in the consideration set. This possibility would be interesting but not particularly new, given the literature on loss and coping. The explanation we propose is not driven by how emotionally involved participants are with a given ethical issue but rather by differential compatibility between ethical attributes and response modes. To address this possible affect-based explanation, in Studies 1 and 2, we provide statistical evidence that affect does not drive our effects. In Study 3, we provide direct evidence that consumers use the principle of compatibility when evaluating an ethical decision. Finally, in Study 4, we demonstrate that consumers give greater weight to ethical attributes in exclusion than inclusion in a laboratory study involving real (versus hypothetical) consideration set formation.

**STUDY 1**

The objective of this study was to demonstrate that consumers weight ethical attribute information more when they form a consideration set by exclusion than by inclusion. The study employed a 2 (task: include versus exclude) × 2 (ethical implications of labor issue: high versus low) between-subjects design. Participants assigned a set of automobiles to either an include or an exclude set, and these judgments enabled us to identify attribute weights using conjoint analysis. There were three automobile attributes, with three levels each, resulting in a complete matrix of 27 automobiles to assign per participant. In our stimuli, we used the term “cars” and therefore use this more informal terminology in the discussion of our results.

**Stimuli and Procedure**

The usable sample for Study 1 consisted of 421 undergraduate students at a large university (261 women, 157 men, and 3 nonresponses) who participated in the study in exchange for course extra credit. We removed 22 participants whose responses indicated that they did not understand the study instructions and procedure.1

Participants were told that they would be sharing their opinions about a variety of cars that differed on three attributes: price, performance, and an ethical labor attribute. They were told that the cars did not differ on anything other than these attributes. The price and performance attributes used standard wording:

**Price:** The final negotiated cost of the car.

**Performance:** Performance ratings for the car from a leading consumer magazine. The performance ratings range from one to ten, with ten being the highest.

The labor practices variable had two levels (manipulated between subjects). Some participants saw a labor attribute with significant ethical implications:

**Labor practices of the manufacturer of the car:**

The manufacturers differ in their treatment of their workers. The best measure of this treatment is the number of lawsuits brought by employees against the management.

Others saw one with relatively weaker ethical implications:

**Labor practices of the manufacturer of the car:**

How much (paid) vacation time the company allows its employees to take per year.

In the case of the lawsuit version of the ethical labor attribute, we also made it clear that the lawsuits pertained to the treatment of the workers only and did not reflect the quality of the cars. After participants read these descriptions, they were given an envelope containing slips of paper that described the cars that could be formed using the three levels of the three attributes, for a total of 27 cars (one car per slip labeled “A” through “AA”) in both the lawsuit and the vacation-days conditions (for a detailed description of the study’s procedure and a list of all the cars in the vacation-days condition, see the Web Appendix at http://www.marketingpower.com/jmarrpi09). The three levels for the price attribute were $11,977, $13,385, and $14,793. The three levels for the performance attribute (on the ten-point scale we described previously) were 6, 7.75, and 9.5. The three levels for the ethical labor attribute were “many more than average,” “average,” and “less than average” in both the lawsuit and the vacation-days conditions. These categories corresponded to “many complaints, including assault charges,” “five–ten lawsuits per year,” and “one or two lawsuits every few years” in the lawsuit condition and to zero vacation days, four days, and one week in the vacation-days condition.

Participants were then asked to sort the automobiles. Participants in the include condition were asked to put all the cars they would be interested in purchasing into an envelope marked “INTERESTED.” Participants in the exclude condition were asked to put all the cars they would not be interested in purchasing into an envelope marked “NOT INTERESTED.” Participants were told that they could select as many or as few cars as they liked. We gave them as much time as they needed to complete this task (approximately ten minutes on average), and we ran the study in groups of 10–40, in which all members of the group participated in either the include or the exclude task.

After completing the sorting task, participants in the include condition wrote down the letter for all the cars they had placed in their envelope. Participants in the exclude condition were asked to put aside their envelope and to record the letters of the cars they had remaining in front of

---

1For example, their models had slopes of zero for all three attributes, they had negative slopes for all three attributes, they chose only one car for their consideration set, or their conjoint model had an extremely small R-square (less than .2).
them. Thus, both groups of participants recorded the cars in their final consideration set, though the method they used to arrive at their consideration set differed by condition.

Finally, both groups rank-ordered the cars in their consideration set, with a rank of one indicating the car they would be most interested in purchasing. They then answered several questions on a nine-point scale that assessed (1) the ethical significance of the labor issue and (2) their affective response to the issue. The three questions assessing ethical significance provided manipulation checks to determine whether number of lawsuits was viewed as a more significant ethical issue. The questions did not form one construct (Cronbach’s $\alpha = .56$), and thus we present their results separately. We summed the four questions assessing emotional reactions to form an overall index of participants’ emotional reaction to the ethical issue (Cronbach’s $\alpha = .85$) (for a list of the ethical significance and emotion questions, see Appendix A).

Results

**Manipulation check.** Both vacation days and lawsuits were considered ethically charged ($M_{\text{vacation days}} = 6.70$, $M_{\text{lawsuits}} = 7.07$; $F(1, 416) = 3.64, p = .06$), with lawsuits marginally more so. However, number of lawsuits was rated as a more important issue than vacation days ($M_{\text{vacation days}} = 6.30$, $M_{\text{lawsuits}} = 6.97$; $F(1, 417) = 7.83, p < .01$) and as more likely to make a difference in labor policy if consumers made purchase decisions based on it ($M_{\text{vacation days}} = 6.02$, $M_{\text{lawsuits}} = 6.90$; $F(1, 416) = 14.51, p < .0001$).

**Size of consideration set.** Consistent with prior research (Heller, Levin, and Goransson 2002; Yaniv et al. 2002), exclusion produced larger consideration sets than inclusion (mean number of cars in consideration set formed by exclusion = 8.79, mean number of cars in consideration set formed by inclusion = 7.04; $F(1, 417) = 24.40, p < .0001$).

**Attribute-weighting results.** We used a hierarchical analysis to test our hypotheses, first obtaining conjoint weights for each participant and then testing whether the weights differed by the response mode.\(^2\) Recall that we suggested two potential drivers of this result. If compatibility drives the differential weighting, individual emotions would add variance to the model that is not explained by task (i.e., participants would have heterogeneous levels of emotional involvement that might actually obscure the compatibility effect). Statistically, this hypothesis would be supported if adding the summed emotional reaction variable to the model as a covariate strengthened the task effect. Conversely, if emotion drives the effect of task on consideration set formation, the emotion measure should mediate the task effects. In addition to models testing the effect of task type on the weighting of the ethical labor attribute both (1) without the emotion covariate in the model and (2) with the emotion covariate, we examined the rank orders assigned to the cars in participants’ consideration set and the set of rejected cars that were not in participants’ consideration sets to provide more process evidence (we provide the details subsequently).

Our primary prediction was that task type (inclusion versus exclusion) would have a main effect on the weighting of the ethical labor attribute such that the ethical attribute would be weighted more in the exclude condition than in the include condition. Our results (running the model without the emotion covariate) support this prediction; the ethical labor attribute was indeed weighted more in the exclude ($M_{\text{slope}} = .33$) than in the include ($M_{\text{slope}} = .27$) condition ($F(1, 417) = 3.71, p = .05$). Adding the emotion covariate to the model strengthened the main effect of task type, in support of the compatibility explanation for our results ($F(1, 412) = 6.67, p = .01$). The R-square for the model with the emotion covariate is .20, compared with an R-square value of .05 for the model without the covariate. Furthermore, the average score on the summed emotion covariate did not differ by task ($F(1, 412) = 1.58, p > .20$), indicating that emotion is not a mediator of the effect. We conducted the subsequent attribute-weighting analyses with the emotion covariate included in the model.

The ethical labor attribute could pertain to either vacation days or lawsuits. As Figure 1 shows, there was a main effect of the type of labor attribute: the labor attribute was weighted more in the lawsuit condition ($M_{\text{slope}} = .34$) than in the vacation-days condition ($M_{\text{slope}} = .26$; $F(1, 412) = 12.29, p < .001$). In addition, the type of labor attribute (lawsuit versus vacation days) interacted with task type (include versus exclude) to influence the weight placed on the ethical attribute ($F(1, 412) = 3.15, p = .08$). Follow-up analyses focusing on the lawsuit and vacation-days conditions, respectively (using the spotlight method proposed by Irwin and McClelland [2001]), revealed that inclusion versus exclusion only had an effect on the weighting of the ethical labor attribute in the lawsuit condition (i.e., only the more important ethical attribute was weighted more heavily in exclusion than in inclusion) ($F(1, 412) = 8.26, p < .01$). Task type did not have a significant effect on ethical attribute weighting in the vacation-days condition ($F(1, 412) = .21, p = .65$). These results suggest a boundary condition for the ethical attribute/compatibility effect we propose; task type (inclusion versus exclusion) has an effect on the

---

\(^2\)We converted negative slopes to zeros for the second stage of the analysis in both Studies 1 and 2.
weighting of ethical attributes only for relatively important ethical attributes.

As a control, we also tested whether there was a main effect of include/exclude on the price or performance slopes. There was not; the weights for price (M_{exclude} = .15, M_{include} = .16) and performance (M_{exclude} = .61, M_{include} = .59) did not differ significantly by task type (F = 1.9 and 1.24, respectively). On average, the weights for the ethical labor attributes were less than the performance weights but higher than the price weights, showing that our effects were not driven by the importance of the ethical attribute.

There is another alternative explanation for our results, apart from the potential role of emotion or attribute importance. Perhaps in exclusion modes, consumers neglect to exclude the more ethical alternatives, but they have no intention of actually choosing them. In other words, perhaps the ethical alternatives are not truly in the consideration set but rather are “tacked on” to help the consumer feel better. Given that the decision processes used in choosing an alternative from a consideration set can differ from those used to form the set (e.g., Chakravarti, Janiszewski, and Ülkümen 2006), perhaps ethicality ceases to matter after the set is formed.

Analysis of rank orders. To test this alternative explanation, we turned to the rank orders of the items in the consideration set. The unit of analysis for this test was the car: We regressed the rank of the car on its average ethicality level across participants, within the consideration set. Ranks could theoretically range from 1 to 27 (the number of possible cars in the consideration set), though, on average, participants had fewer cars in their consideration sets (approximately 7 in include and 9 in exclude). Regressing participants’ ranks on the mean ethicality levels showed whether the more ethical cars were simply added to the consideration set but not actually considered—that is, whether the cars with the best level on the ethical labor attribute (the most vacation days per worker or the least lawsuits brought against management) had the worst ranks. Instead, we found that in both conditions, rank significantly reflected ethicality ((t(1, 36) = 3.70, p < .001), such that more ethical cars had higher ranks. In addition, the relationship between ranking and ethicality did not depend on include/exclude (F(1, 36) = .92, p = .36). Thus, it is not the case that in the exclude condition the most ethical cars were in the consideration set but were consistently ranked last.

Rejected product alternatives. Next, we turned our attention to examining the rejected cars—that is, the cars that were not in participants’ consideration sets. These cars provide more evidence for the processes our decision makers used to form their consideration set. Our compatibility explanation suggests that it is easier to leave out the more ethical products in include than exclude response modes and, conversely, to reject explicitly the unethical products in exclude versus include. To test this prediction, we compared the percentage of rejected cars with the best, worst, and average rating on the ethical attribute, respectively, across task type. As Figure 2 shows, the percentage of rejected cars with the best rating on the ethical attribute was greater in the include (27.4%) than in the exclude (25.8%) condition (F(1, 409) = 9.33, p < .001). Conversely, the percentage of rejected cars with the worst rating on the ethical attribute (43.0%) was greater in the exclude than in the include (40.0%) condition (F(1, 409) = 12.77, p < .001).

Overall, there were significant linear (F(1, 412) = 774.95, p < .0001) and quadratic (F(1, 412) = 30.25, p < .0001) trends in these data, but only the linear trend differed significantly by the inclusion versus exclusion task (F(1, 412) = 5.40, p < .05).

Discussion

Study 1 provided evidence of the task goal compatibility explanation we propose. Consistent with our expectations, participants weighted ethical attributes more in exclusion than in inclusion modes. There were no differences in weight for the other attributes (i.e., price and performance). Study 1 also ruled out an affect-based explanation for our task-type results. However, we must still address an alternative explanation to establish that our results are driven by compatibility. This explanation is based on Shafir’s (1993) finding that positive attributes loom larger in choice tasks and negative attributes loom larger in rejection tasks. Because participants may have viewed our ethical attribute (potentially harmful labor practices) negatively, Study 2 replicates the results of Study 1 and includes a control for ethical attribute frame.

**STUDY 2**

The primary objective of Study 2 was to establish that Study 1’s findings were not attributable to framing. Study 2 had a 2 (task: include versus exclude) × 2 (ethical attribute frame: include versus exclude) design. Figure 2 shows the percentage of rejected cars with the best, worst, and average rating on the ethical attribute, respectively, across task type. As Figure 2 shows, the percentage of rejected cars with the best rating on the ethical attribute was greater in the include (27.4%) than in the exclude (25.8%) condition (F(1, 409) = 9.33, p < .001). Conversely, the percentage of rejected cars with the worst rating on the ethical attribute (43.0%) was greater in the exclude than in the include (40.0%) condition (F(1, 409) = 12.77, p < .001).

Overall, there were significant linear (F(1, 412) = 774.95, p < .0001) and quadratic (F(1, 412) = 30.25, p < .0001) trends in these data, but only the linear trend differed significantly by the inclusion versus exclusion task (F(1, 412) = 5.40, p < .05).

Table 2 shows the percentage of rejected cars with the best, worst, and average rating on the ethical attribute, respectively, across task type. As Figure 2 shows, the percentage of rejected cars with the best rating on the ethical attribute was greater in the include (27.4%) than in the exclude (25.8%) condition (F(1, 409) = 9.33, p < .001). Conversely, the percentage of rejected cars with the worst rating on the ethical attribute (43.0%) was greater in the exclude than in the include (40.0%) condition (F(1, 409) = 12.77, p < .001).

Overall, there were significant linear (F(1, 412) = 774.95, p < .0001) and quadratic (F(1, 412) = 30.25, p < .0001) trends in these data, but only the linear trend differed significantly by the inclusion versus exclusion task (F(1, 412) = 5.40, p < .05).

Discussion

Study 1 provided evidence of the task goal compatibility explanation we propose. Consistent with our expectations, participants weighted ethical attributes more in exclusion than in inclusion modes. There were no differences in weight for the other attributes (i.e., price and performance). Study 1 also ruled out an affect-based explanation for our task-type results. However, we must still address an alternative explanation to establish that our results are driven by compatibility. This explanation is based on Shafir’s (1993) finding that positive attributes loom larger in choice tasks and negative attributes loom larger in rejection tasks. Because participants may have viewed our ethical attribute (potentially harmful labor practices) negatively, Study 2 replicates the results of Study 1 and includes a control for ethical attribute frame.

**STUDY 2**

The primary objective of Study 2 was to establish that Study 1’s findings were not attributable to framing. Study 2 had a 2 (task: include versus exclude) × 2 (ethical attribute frame: include versus exclude) design. Figure 2 shows the percentage of rejected cars with the best, worst, and average rating on the ethical attribute, respectively, across task type. As Figure 2 shows, the percentage of rejected cars with the best rating on the ethical attribute was greater in the include (27.4%) than in the exclude (25.8%) condition (F(1, 409) = 9.33, p < .001). Conversely, the percentage of rejected cars with the worst rating on the ethical attribute (43.0%) was greater in the exclude than in the include (40.0%) condition (F(1, 409) = 12.77, p < .001).

Overall, there were significant linear (F(1, 412) = 774.95, p < .0001) and quadratic (F(1, 412) = 30.25, p < .0001) trends in these data, but only the linear trend differed significantly by the inclusion versus exclusion task (F(1, 412) = 5.40, p < .05).

Discussion

Study 1 provided evidence of the task goal compatibility explanation we propose. Consistent with our expectations, participants weighted ethical attributes more in exclusion than in inclusion modes. There were no differences in weight for the other attributes (i.e., price and performance). Study 1 also ruled out an affect-based explanation for our task-type results. However, we must still address an alternative explanation to establish that our results are driven by compatibility. This explanation is based on Shafir’s (1993) finding that positive attributes loom larger in choice tasks and negative attributes loom larger in rejection tasks. Because participants may have viewed our ethical attribute (potentially harmful labor practices) negatively, Study 2 replicates the results of Study 1 and includes a control for ethical attribute frame.
frame: positive versus negative) design. Although we expected to replicate the framing effects found in previous research because exclusion is more compatible with negatively framed attributes (and vice versa), we also aimed to show that our task-type effects could be obtained beyond the effect of frame.

**Stimuli and Procedure**

Three hundred forty-six undergraduate students at a large university participated in this study for course extra credit. Following the guidelines established in Study 1, we removed 25 participants whose responses indicated that they did not understand the study procedure. The remaining sample consisted of 321 participants (168 women, 148 men, and 5 nonrespondents). The design of the study was similar to that of Study 1, except for the ethical labor attribute. In Study 2, the ethical labor attribute was management’s treatment of workers, which was framed either positively or negatively. In the positive frame, the ethical labor attribute was discussed in terms of the number of lawsuits against management that had been successfully resolved:

**Labor practices of the manufacturer of the car:** The best measure of the labor practices of the manufacturer is the percentage of lawsuits brought by employees against the management that have been successfully resolved. A lawsuit that has been resolved is one that has been resolved in favor of the workers, that is, one that has fully addressed the issues the workers brought up in the lawsuit. The issues range from unfair firing to sexual harassment to maternity leave issues. Judgments of whether the lawsuit was resolved are made by an independent organization.

The manufacturers differ in the number of lawsuits that they have been able to resolve. This difference is reflected in the percentage of total suits brought against the company by its workers that have been successfully resolved. Percentages range from 0%, meaning none of the suits have been successfully resolved by management, to 100%, meaning all the suits have been successfully resolved. Obviously, 100% is the best for workers and 0% is the worst.

In the negative frame, the ethical labor attribute was discussed in terms of the number of lawsuits unresolved:

The manufacturers differ in the number of unresolved lawsuits. This difference is reflected in the percentage of total suits brought against the company by its workers that management has not resolved. Percentages range from 0%, meaning no lawsuits remain unresolved, to 100%, meaning all the lawsuits against management remain unresolved. Obviously, 0% is the best for workers and 100% is the worst.

The three attribute levels for the positive frame were 90%, 50%, and 20% of lawsuits resolved. For the negative frame, the levels were the analogous 10%, 50%, and 80% of lawsuits unresolved. Price and performance were identical in description and attribute level to Study 1. We combined the three levels of the three attributes to form 27 distinct cars in both conditions, as in Study 1.

The procedure for the study was analogous to Study 1, with participants first completing the sorting and then the recording and ranking tasks. We also included the emotion questions from Appendix A (Cronbach’s α = .82).

**Results**

**Impact of task type and ethical attribute framing on weighting of ethical attributes.** As in Study 1, there was a main effect of task type on the ethical slope: The ethical attribute was weighted more in the exclude condition (M_{slope} = .32) than in the include condition (M_{slope} = .26), both when the emotion covariate was included in the model (F(1, 315) = 4.05, p < .05) and when it was not (F(1, 318) = 3.22, p = .07). As in Study 1, these results provide further evidence that our results are driven by compatibility rather than emotional reactions and/or coping. As in Study 1, there were no main effects of task type on weighting of price (M_{exclude slope} = .15, M_{include slope} = .15) or performance (M_{exclude slope} = .56, M_{include slope} = .59; Fs = .04 and .26, respectively).

However, the most important analysis for this study was whether task type exerted an effect regardless of frame. There was an overall effect of frame on the ethical attribute slope (F(1, 318) = 4.27, p < .05, β = -.03), but this effect was not related to the task-type effect. Task type was significant when we controlled for frame, and there was no frame × task-type interaction (F = .97).

**Replication of Study 1 results.** Study 2 also provided us an opportunity to replicate the set-size and rank-order analyses from Study 1. Consistent with our prior results and previous research, the consideration sets formed by excluding were larger (M = 8.39) than those formed by including (M = 7.28; F(1, 301) = 5.59, p < .05). Again, this effect was not due to participants’ tacking on the ethical alternatives in exclusion. As in Study 1, cars with worse levels on the ethical attribute were given worse ranks in the exclude condition (t(1, 33) = 3.34, p < .01). However, this result was not replicated in the include condition, probably because of power issues (there were fewer cars in the consideration sets formed using inclusion). We also found that the percentage of rejected cars with the worst rating on the ethical attribute was significantly greater in the exclude (41%) than in the include (39%) condition (F(1, 295) = 5.69, p < .05).

**Discussion**

Study 2 provided further evidence for the task goal compatibility explanation we propose for the differential weighting of ethical attributes in inclusion versus exclusion by ruling out framing as an alternative explanation. Regardless of whether an ethical attribute is framed positively or negatively, it is weighted more in exclusion modes than in inclusion modes. Study 2 also replicated the ranking results from Study 1.

Our aim in Study 3 was to provide a more direct view of the differential compatibility of ethical attributes with exclusion and inclusion modes. Although Studies 1 and 2 provided evidence that consumers’ behavior is consistent with our explanation, they did not test whether a particular decision behavior seems more compatible to consumers than other behaviors. For example, do consumers believe that not including ethical alternatives in a consideration set is a less direct statement of ethicality than explicitly excluding ethical alternatives? We designed Study 3 to provide this evidence.

**STUDY 3**

Is excluding ethical alternatives from consideration less justifiable than the normative equivalent (i.e., neglecting to
include the alternatives? If we can show that greater weighting of ethical attributes feels more “right” in inclusion than in inclusion, we will have powerful evidence for our compatibility explanation. Thus, Study 3 has a completely different methodology from the other two studies; in Study 3, we ask participants to judge the behavior of another consumer.

We presented participants with one of four decisions purportedly made by a consumer in the market for a new car. We asked participants to judge the decision (by responding to the seven questions listed in Appendix B) to assess whether they thought the decision was a good decision, whether it was justifiable, and so on. The decisions differed by task and whether they were consistent with the compatibility explanation we propose. The study had a 2 (task: include versus exclude) × 2 (cars included in final consideration set: cars with best level on ethical attribute only versus cars with average or worst level on ethical attribute only) design. Note that because one set (the set with only the cars with the best level on the ethical attribute) is smaller than the other set, we can control for participants’ expectations about consideration set size in the two modes. Indeed, our expected results—that the more ethical, smaller consideration set will be viewed as more justifiable when formed by exclusion—run counter to the usual exclusion = larger set size finding.

Stimuli and Procedure

We recruited 273 (153 women, 117 men, and 3 nonrespondents) native English speakers by e-mail to participate in this study online. Participants ranged in age from 20 to 68 years. The average age was 33.8. We asked all participants to evaluate a decision made by Steve, a hypothetical consumer in the market for a new car. The scenario stated that Steve had identified six cars to choose from that all had the same price. The cars differed only on the number of lawsuits filed against management and the performance of the car, which were described in the same manner as in Study 1. Table 1 lists the six cars; the same information was presented to participants in a similar table.

Participants were told either that Steve had decided to exclude a subset of these six cars from further consideration or that he had decided to include a subset for consideration. Participants then saw Steve’s final consideration set, which had two or four cars—either the two cars with the best level on the ethical attribute or the four cars with the average and worst level on the ethical attribute. Table 2 summarizes these conditions. Note that Conditions 1 and 2 are normatively equivalent, as are Conditions 3 and 4.

Table 1

<table>
<thead>
<tr>
<th>Car</th>
<th>Performance</th>
<th>Lawsuits Brought Against Car Manufacturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>9.50</td>
<td>Many</td>
</tr>
<tr>
<td>B</td>
<td>7.75</td>
<td>Many</td>
</tr>
<tr>
<td>C</td>
<td>9.50</td>
<td>Average</td>
</tr>
<tr>
<td>D</td>
<td>7.75</td>
<td>Average</td>
</tr>
<tr>
<td>E</td>
<td>9.50</td>
<td>Few</td>
</tr>
<tr>
<td>F</td>
<td>7.75</td>
<td>Few</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th>Study 3: Consideration Set Composition and Task-Type Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Consideration Set</td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>Include</td>
</tr>
<tr>
<td>Exclude</td>
</tr>
</tbody>
</table>

After participants read about Steve’s decision, they were asked to evaluate it. Appendix B shows the seven questions on which participants rated the decision (1 = “definitely not,” and 9 = “definitely so”). We expected all the questions to reflect the same underlying factor (whether the decision made was considered a good one), and we expected these ratings to be higher for decisions consistent with our compatibility explanation. That is, when including, participants should find it more acceptable to have a final consideration set that includes alternatives that do not have the best rating on the ethical attribute, but when excluding, participants should find it more acceptable to have a final consideration set that includes only the alternatives with the best level on the ethical attribute. Such a result would provide evidence that consumers find explicit rejection of ethical alternatives to be less permissible than simply not including the ethical alternatives in the set.

Results

Although we expected all seven of the decision ratings questions to reflect the same underlying factor (whether the decision was a good decision), the Cronbach’s alpha for all seven ratings was unacceptably low (α = .79). An exploratory factor analysis showed two factors: the factor we expected, on which all but two of the seven questions loaded, and a second factor, which consisted of the “quick” and “easy” questions (“Do you think Steve’s decision was easy for him to make?” “Was Steve able to make his decision quickly?”). These factors had eigenvalues of 3.29 and 1.52, respectively.

A 2 (task type) × 2 (final consideration set) factor analysis of variance using the two factors (termed “good decision” and “ease”) as the dependent variables enabled us to test our hypotheses. There was an overall main effect of consideration set for both the good decision and the ease factors. The consideration set that included only the two cars with the highest level on the ethical attribute was rated as a significantly better decision (F(1, 269) = 150.42, p < .0001) and a marginally easier one (F(1, 268) = 3.16, p = .08). In some ways, these effects are manipulation checks; participants noticed that these consideration sets were more ethical and, presumably, easier to form (given that they contained two versus four cars and the dominating car with the best rating on both the ethical attribute and performance).

There were no main effects for task type for either factor, showing that inclusion (or exclusion) is not considered a more justifiable or easier mode overall. For marketers, this is good news because it suggests that consumers could be
encouraged toward one mode or another. However, our primary interest was in the interaction between consideration set and task type. We expected the more ethical consideration set (i.e., the set that included only the cars with the best level on the ethical attribute) to be considered especially justifiable in exclusion versus inclusion, and vice versa. As Figure 3 shows, we obtained this result (F(1, 269) = 4.00, p < .05). Although there was no difference in how the decision was rated by task type for the less ethical consideration set (M_{exclude} = 3.83, M_{include} = 4.10; F = 1.21, p = .27), the interaction was driven by the difference between how the decision was rated for the more ethical consideration set, such that it is considered a better decision in exclusion than in inclusion (M_{exclude} = 6.64, M_{include} = 6.12; F(1, 269) = 3.21, p < .07). The interaction between consideration set and task type was not significant for the ease factor, with average ratings near the midpoint of the scale (approximately six) for all conditions (F = 1.74, p = .19).

Discussion

Study 3 provided evidence that consumers believe that excluding ethical alternatives from consideration is less justifiable than the normative equivalent (i.e., neglecting to include the same alternatives). In conjunction with the results of Studies 1 and 2, the results of Study 3 provide further support for our proposition that the weighting of ethical attributes is more compatible with exclusion than with inclusion. Participants’ ratings of the quality, justifiability, and so on, of Steve’s decision suggest that explicitly rejecting ethical alternatives is associated with rejecting the ethical principle they reflect, and most consumers are, at least explicitly, unwilling to do this. However, inclusion does not highlight the ethical attribute in the same manner, and consumers consider it acceptable simply not to include ethical alternatives, because this action does not imply that one is rejecting an ethical principle.

Figure 3

STUDY 3: JUDGMENTS OF THE JUSTIFIABILITY OF CONSIDERATION SET

![Graph showing judged justifiability of consideration set](image)

Cars in Hypothetical Shopper Steve’s Consideration Set

- Consideration set formed using exclusion
- Consideration set formed using inclusion

STUDY 4

Studies 1–3 used hypothetical scenarios. Study 4 replicates our primary result (i.e., that ethical attributes receive greater weight when consideration sets are formed using exclusion than inclusion) in a task with real consequences. It also uses another product category, shampoo. There was one between-subject variable, include versus exclude, and participants formed actual consideration sets using bottles of shampoo that were either tested on animals or not. At the conclusion of the study, participants received a full-sized bottle of the shampoo from their consideration set that they ranked as their number-one choice.

Pretest and Stimuli

We ran a computer-based pretest to select the shampoos for the study. Our goal was to do as well as we could (given limitations in the marketplace) at equating the set of “ethical” versus “less ethical” shampoos on other relevant attributes, such as preference and price, to serve as an analogy to the orthogonal conjoint matrix in the first two studies. We asked 31 undergraduate students to rate 31 shampoos (15 that were not tested on animals and 16 that were) on two nine-point rating scales (how much they liked each shampoo and how likely they were to purchase each). We showed them large photos of the shampoos and provided the retail price but did not tell them which shampoos were tested on animals.

To ensure that the price and purchase likelihood judgments were equated on attributes other than animal testing, we then selected eight shampoos that were not tested on animals (M_{price} = $8.09, M_{purchase likelihood} = 3.54, M_{liking} = 4.29), and eight shampoos that were tested on animals (M_{price} = $8.05, M_{purchase likelihood} = 3.40, M_{liking} = 4.19).

We also asked pretest participants to rate whether they thought the shampoo brands were tested on animals (1 = “definitely not,” 5 = “not sure,” and 9 = “definitely are”); there was not a significant difference between the ratings for the brands we selected to make up the two groups (for shampoos not tested on animals, M_{no test} = 4.27; for shampoos tested on animals, M_{test} = 4.57; F = 1.74, p = .20). Table 3 provides detailed information on the 16 shampoos that made up these two groups.

Procedure

The sample for Study 4 was 28 doctoral students and staff at a large university (16 women, 12 men) who participated in the study in exchange for a free bottle of shampoo. Participants ranged in age from 23 to 58 years, with an average age of 35.5. We told them that they would be participating in a study designed to understand how people shop for shampoos and that the 16 shampoos they would “shop” for differed on three main attributes: whether the shampoo was for moisturizing hair, the retail price, and whether the shampoo’s manufacturer tests its product on animals. To explain the ethical animal-testing attribute, we told them the following:

Some shampoo manufacturers test their products on animals and some do not. Shampoos with a “Cruelty

4We obtained shampoo prices from http://shopping.msn.com/, a comparison shopping site that lists prices for each shampoo from several Web sites, including Drugstore.com, Amazon.com, Haircarechoices.com, and Beautydepot.com.
We predicted a main effect of task type on the weighting of the ethical animal-testing attribute, and our results support this prediction. The ethical animal-testing attribute was weighted more in the exclude than in the include task. As in Studies 1 and 2, we first obtained an attribute weight for animal testing for each participant. Unlike in the conjoint studies, however, which had the benefit of orthogonal matrices, we needed to add some controls to isolate the animal-testing attribute properly. Thus, we included the animal-testing variable in this first stage of our analysis, as well as the shampoos’ price per ounce and shampoo type (whether the shampoo was for moisturizing hair, determined by reading the shampoo labels). Then, we regressed each participant’s slope for animal testing on task type (include versus exclude), with the summed emotional reaction variable ($\alpha = .91$) included in the model as a covariate.

We predicted a main effect of task type on the weighting of the ethical animal-testing attribute, and our results support this prediction. The ethical animal-testing attribute was weighted more in the exclude ($M_{\text{slope}} = .29$) than in the include ($M_{\text{slope}} = .11$) condition ($F(1, 25) = 5.02, p < .05$).

Discussion

The results of Study 4 provided evidence that our compatibility effect holds (1) across a new product category (shampoos) and (2) with a new ethical attribute (the testing of consumer products on animals). Most important, however, Study 4 also demonstrated the power of the ethical attribute/exclusion compatibility effect by demonstrating it in a context in which participants formed consideration sets with real monetary and usage consequences.

GENERAL DISCUSSION

Why are ethical products not more popular? No one set of studies can answer this question, but this research

Table 3
STUDY 4: SHAMPOO BRANDS

<table>
<thead>
<tr>
<th>Shampoo Brand</th>
<th>Name</th>
<th>Cruelty Free</th>
<th>Moisturizing</th>
<th>Ounces</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burt’s Bees</td>
<td>More Moisture Raspberry and Brazil Nut Shampoo</td>
<td>Yes</td>
<td>Yes</td>
<td>12</td>
<td>$7.99</td>
</tr>
<tr>
<td>Neutrogena</td>
<td>Triple Moisture Cream Lather Shampoo</td>
<td>No</td>
<td>Yes</td>
<td>8.5</td>
<td>$5.99</td>
</tr>
<tr>
<td>Healthy Sexy Hair</td>
<td>Soy Milk Shampoo</td>
<td>No</td>
<td>Yes</td>
<td>13.5</td>
<td>$9.78</td>
</tr>
<tr>
<td>Tigi</td>
<td>Moisture Maniac Shampoo</td>
<td>No</td>
<td>Yes</td>
<td>12</td>
<td>$9.98</td>
</tr>
<tr>
<td>Alba Hawaiian</td>
<td>Honeydew Nourishing Hair Wash</td>
<td>Yes</td>
<td>Yes</td>
<td>12</td>
<td>$7.99</td>
</tr>
<tr>
<td>Aveda</td>
<td>Shampure</td>
<td>Yes</td>
<td>No</td>
<td>8.5</td>
<td>$9.00</td>
</tr>
<tr>
<td>Paul Mitchell</td>
<td>Shampoo One</td>
<td>Yes</td>
<td>No</td>
<td>10.14</td>
<td>$8.19</td>
</tr>
<tr>
<td>Burt’s Bees</td>
<td>Super Shiny Grapefruit &amp; Sugar Beet Shampoo</td>
<td>Yes</td>
<td>No</td>
<td>12</td>
<td>$7.20</td>
</tr>
<tr>
<td>Neutrogena</td>
<td>Clean Replenishing Shampoo</td>
<td>No</td>
<td>No</td>
<td>10.1</td>
<td>$4.69</td>
</tr>
<tr>
<td>Infusium 23</td>
<td>(Repairologie Shampoo)</td>
<td>No</td>
<td>Yes</td>
<td>16</td>
<td>$6.99</td>
</tr>
<tr>
<td>Alba Botanicals</td>
<td>Balancing Shampoo</td>
<td>Yes</td>
<td>No</td>
<td>12</td>
<td>$7.02</td>
</tr>
<tr>
<td>Big Sexy Hair</td>
<td>Big Volume Shampoo</td>
<td>No</td>
<td>No</td>
<td>13.5</td>
<td>$9.99</td>
</tr>
<tr>
<td>Tigi</td>
<td>Control Freak Shampoo</td>
<td>No</td>
<td>No</td>
<td>12</td>
<td>$9.98</td>
</tr>
<tr>
<td>Infusium 23</td>
<td>(Moisturologie Shampoo)</td>
<td>No</td>
<td>Yes</td>
<td>16</td>
<td>$6.99</td>
</tr>
<tr>
<td>Paul Mitchell</td>
<td>Super Skinny Daily Shampoo</td>
<td>Yes</td>
<td>No</td>
<td>10.14</td>
<td>$8.34</td>
</tr>
<tr>
<td>Aveda</td>
<td>Rosemary Mint Shampoo</td>
<td>Yes</td>
<td>No</td>
<td>8.5</td>
<td>$9.00</td>
</tr>
</tbody>
</table>

Free” sticker on their bottle have verified that they DO NOT conduct or commission any animal tests not required by law on ingredients, formulations, or finished products and that they pledge not to do so in the future. Any shampoo without the “Cruelty Free” sticker manufactures products that ARE tested on animals.

We obtained this description of animal testing and the data on companies that do and do not conduct animal tests from www.caringconsumer.com, a Web site sponsored by PETA (2007).

Rather than working with slips of paper with attribute information written on them, as in Studies 1 and 2, participants evaluated actual bottles of shampoo in a mock store environment. The bottles had labels with price and “cruelty free” (if relevant) placed on them along with all the other descriptions already on the bottles. The shampoo bottles were lined up, as if in a retail display, always in the same order in which they are listed in Table 3. Participants entered the area one at a time and were told that they would be given a $10.00 shopping budget and that with that budget they would “buy” a bottle of shampoo. If the shampoo cost less than $10.00, they would receive the change remaining from their shopping budget along with the shampoo. (This procedure enabled us to account for variations in price among the shampoos). The average price of shampoo chosen was $7.73.

When participants were shown the shampoo display, we told them that 16 shampoos constituted too big a set to evaluate carefully, so we were going to let them narrow the set. In the include condition, we gave them a bin labeled “INTERESTED” and asked them to put any shampoos that interested them in the bin, so that they could then spend as much time as they wanted with just those shampoos. In the exclude condition, the bin was labeled “NOT INTERESTED,” and the participants were told they could put the shampoos that did not interest them in the bin and then spend as much time as they liked with the remaining shampoos. Participants were then allowed to spend as much time examining and sorting the shampoos as they needed; on average, the task took approximately seven minutes.

After completing the sorting task, participants ranked the shampoos in their consideration set (i.e., either the shampoos in the “INTERESTED” bin or the ones not in the “NOT INTERESTED” bin, depending on condition). At this point, we reminded participants that they would receive their top-ranked shampoo and any change remaining from their $10.00 shopping budget. Then, participants answered the same ethical significance and emotion questions used in Study 1 (rephrased to ask about animal testing instead of the ethical labor issues from Studies 1 and 2). After completing the survey, participants “purchased” their chosen shampoo.

Results

Our primary hypothesis was that animal testing would be weighted more in the exclude than in the include task. As in Studies 1 and 2, we first obtained an attribute weight for animal testing for each participant. Unlike in the conjoint studies, however, which had the benefit of orthogonal matrices, we needed to add some controls to isolate the animal-testing attribute properly. Thus, we included the animal-testing variable in this first stage of our analysis, as well as the shampoos’ price per ounce and shampoo type (whether the shampoo was for moisturizing hair, determined by reading the shampoo labels). Then, we regressed each participant’s slope for animal testing on task type (include versus exclude), with the summed emotional reaction variable ($\alpha = .91$) included in the model as a covariate.

We predicted a main effect of task type on the weighting of the ethical animal-testing attribute, and our results support this prediction. The ethical animal-testing attribute was weighted more in the exclude ($M_{\text{slope}} = .29$) than in the include ($M_{\text{slope}} = .11$) condition ($F(1, 25) = 5.02, p < .05$).
attempts to illuminate one possibility: perhaps some response modes do not naturally lend themselves to the consideration of ethical attributes. In our studies, participants expressed ethical values more in exclusion than in inclusion modes of consideration set formation. When asked to judge the behavior of another decision maker, they made it obvious why: They find that the expression of ethical attributes is more compatible with (i.e., more justifiable and morally relevant in) exclusion than with inclusion modes.

Theoretical and Practical Implications

Unfortunately for marketers hoping to encourage increased consideration of ethical products, marketplace (e.g., retail) consideration set formation is probably often accomplished by inclusion. Although exclusion may be a natural mode of consideration set formation for product categories with smaller numbers of items (Heller, Levin, and Goransson 2002; Ordóñez, Benson, and Beach 1999), inclusion appears to be the default for larger assortments (Heller, Levin, and Goransson 2002; Levin et al. 2001). In today’s crowded marketplace, it is likely that many consumers are naturally including items when forming consideration sets and thus are not expressing their ethical values as much as they might if they were in a more compatible decision mode. The comparative lack of popularity of many ethical products might be partially explained by this notion. Because most retail environments present consumers with a relatively large assortment of products, consumers may default to an include strategy, which does not encourage a focus on ethical attributes.

Unlike purchasing, public policy and political behavior seem to operate more in an exclusion response mode. It is common practice for consumer groups and even governments to call for exclusionary behavior toward companies and countries when they behave unethically, such as calling for boycotts (Klein, Smith, and John 2004). Protected values seem to find their clearest expression in this type of exclusionary public behavior, perhaps because it is easier to rally supporters with strong messages of exclusion than with messages of inclusion. This difference in the natural modes of purchasing and public policy may account for some preference reversals between consumers’ purchasing and their other behavior (e.g., who they vote for, what petitions they sign, what causes them to contribute money to various causes); further research could explore this possibility.

In the meantime, how might marketers with a vested interest guide consumers toward more ethical products? Our results suggest that encouragement of an exclusion strategy should help ethical products receive greater consideration in the marketplace. In Study 3, participants rated exclusion as just as defensible and easy, overall, as inclusion; thus, pushing consumers toward one or the other mode may not be difficult. For example, salespeople could guide customers toward eliminating alternatives rather than including them. Advertising for ethical products could encourage consumers not to “leave us out” or could show people making decisions using exclusion as a model. Giving consumers fewer alternatives to choose from may also encourage the use of an exclusion strategy. Marketers could also choose to follow the example of the many existing companies and organizations that simply provide consumers with a set of prescreened alternatives from which to choose (e.g., fair-trade stores, shopping guides), thus doing the exclusion work for the consumer.

Limitations and Further Research

Our work focuses on consideration set formation, a stage we believe is particularly important for ethical attributes and marketers. Although consideration sets help determine choice (e.g., Hauser 1978), the two stages are not always influenced by identical factors (Chakravarti, Janiszewski, and Ulkkümen 2006; Erdem and Swait 2004; Van Zee, Paluchowski, and Beach 1992). The conjoint framework, though ideal for measuring consideration set weights, is not appropriate for measuring choice, because we used a standard full matrix (with a dominating alternative). We assume that more ethical consideration sets logically lead to more ethical ultimate choices but hope that further research can confirm this assumption.

CONCLUSION

Although there has been some prior work on compatibility and ethical attributes (Irwin 1994; Irwin and Baron 2001; Irwin et al. 1993; Irwin and Spira 1997; Sunstein et al. 2002; Tenbrunsel et al. 2000), the area is still rather new. Instead of focusing only on the ways consumers make mistakes in their judgments by not manifesting their values in their purchasing behavior, further research should (1) uncover when consumers do express their ethical values and (2) address why typical purchasing situations may not be as compatible with these expressions as we might like.

APPENDIX A

Ethical Significance Questions

• In your opinion, how ethically charged is the labor issue of amount of paid vacation time for workers per year [management treatment of workers]?
• How important is the issue of amount of paid vacation time for workers per year [management treatment of workers] to you personally?
• To what extent do you think it would make a difference (i.e., a change in labor policy) if people used the issue of amount of paid vacation time for workers per year [management treatment of workers] to guide their purchasing?

Emotion Questions

• How angry does the thought of little or no paid vacation time for workers per year [poor management treatment of workers] make you?
• How sad does the thought of little or no paid vacation time for workers per year [poor management treatment of workers] make you?
• How guilty would you have felt, when you were picking which cars you would not have considered buying (which cars to exclude) [which cars to consider buying (which cars to include)], if you had ignored amount of paid vacation time for workers per year [management treatment of workers]? That is, how guilty would you feel picking which cars to exclude [include] by price and performance completely, ignoring vacation time for workers [management treatment of workers]?
• How happy would you be to get the car whose workers got the most paid vacation time per year [whose manufacturer had the least lawsuits brought against it]?
APPENDIX B

Study 3: Decision Evaluation Questions and Factor Loadings

1. Is Steve doing what many people would do in the same situation? (good decision = .60, ease = .18)
2. Is Steve a good decision maker? (good decision = .91, ease = .03)
3. Do you think Steve’s decision was easy for him to make? (good decision = .13, ease = .89)
4. Was Steve able to make his decision quickly? (good decision = .04, ease = .90)
5. Did Steve make the right decision, in your opinion? (good decision = .90, ease = .09)
6. Could Steve justify his decision to someone else? (good decision = .80, ease = .05)
7. Is Steve a good person? (good decision = .73, ease = .04)

REFERENCES


APPENDIX B
Study 3: Decision Evaluation Questions and Factor Loadings

1. Is Steve doing what many people would do in the same situation? (good decision = .60, ease = .18)
2. Is Steve a good decision maker? (good decision = .91, ease = .03)
3. Do you think Steve’s decision was easy for him to make? (good decision = .13, ease = .89)
4. Was Steve able to make his decision quickly? (good decision = .04, ease = .90)
5. Did Steve make the right decision, in your opinion? (good decision = .90, ease = .09)
6. Could Steve justify his decision to someone else? (good decision = .80, ease = .05)
7. Is Steve a good person? (good decision = .73, ease = .04)

REFERENCES


