Rejecting a Bad Option Feels like Choosing a Good One

Hannah Perfecto[1]

Jeff Galak[2]

Joseph P. Simmons[3]

Leif D. Nelson[1]

[1]University of California, Berkeley

[2]Carnegie Mellon University

[3]University of Pennsylvania

Author Note:

Hannah Perfecto and Leif D. Nelson, Haas School of Business, University of California, Berkeley; Jeff Galak, Tepper School of Business, Carnegie Mellon University; Joseph P. Simmons, The Wharton School, University of Pennsylvania. Parts of this article were presented at the Annual Meeting of the Society of Judgement and Decision Making in 2015 and at the conference for the Society of Consumer Psychology in 2016. This research was conducted as part of Hannah Perfecto's doctoral dissertation. Correspondence concerning this article should be addressed to Hannah Perfecto, Haas School of Business, University of California, Berkeley, 545 Student Services Building #1900, Berkeley, CA 94720; hannah_perfecto@haas.berkeley.edu.

Abstract

The authors propose a novel framing effect, *attribute matching*, whereby matching a salient attribute of the decision frame with that of the decision's options facilitates decision making. In Study 1, participants choosing the more attractive of two faces and participants rejecting the less attractive faces reported greater confidence in and perceived consensus around their decision. Using positive and negative words, Study 2 showed the size of this effect is moderated by the attribute's extremity. Study 3 found decision ease to mediate these changes in confidence and consensus estimates. Consistent with a misattribution account, when participants were alerted to this external source of ease in Study 4, the effect disappeared. The final study (Study 5) extended attribute matching beyond valence to objective judgments of calorie contents..

Keywords: framing, meta-cognition, judgment, matching

Preferences are malleable. People make different choices depending on whether options are framed as gains or losses (Tversky & Kahneman, 1981), how favorably an option compares to other options that just happen to be in front of them (Tversky & Shafir, 1992), the time frame over which an attribute is described (e.g., price per year vs. price per month; Burson, Larrick, & Lynch, 2009), the name given to an option (Read et al., 2005), etc. Based on such findings, there is now a widespread appreciation of the power of "choice architecture," of the fact that how choice options are arrayed and described can exert a powerful influence on the decisions that people make (Thaler & Sunstein, 2008; Kahneman & Tversky, 1984).

In this paper, we describe a simple framing manipulation that affects not what people choose, but rather how they feel about their choice. Notably, we show that the same framing manipulation can, by changing feelings of confidence, also influence people's beliefs about the choices of others. This is an important contribution, as practitioners are not only in the business of altering preferences; they are also in the business of altering how people feel about the preferences they already have, such as when politicians seek to strengthen the attitudes of those who are already inclined to prefer the campaign's candidate, or when marketers seek to strengthen the attitudes of those who are already inclined to prefer the company's product.

Such campaigns exist because two prospective voters might agree in their preference, but the voter confident in her preference will be more likely to actually vote. Indeed, people are more likely to act on behalf of preferences that are confidently held (e.g., Petty & Krosnick, 1995), and that they think others would endorse (Cialdini, Reno, & Kalgren, 1990). On the other hand, people are less likely to procrastinate in making a choice if they are confident as to which choice to make and believe others would decide similarly. For example, a patient deciding which of two medical procedures to undergo is more likely to make that decision quickly if she is confident

about which choice to make, and if she thinks that others would make the same choice. It is therefore important to not only identify interventions that affect what people choose, but to also identify interventions that affect how people feel about those choices.

In thinking about the variables that influence decision confidence and consensus estimation, we start with the most basic of features of a decision: (1) the valence of the options, and (2) how the decision is framed. Option valence has already been shown to have some unexpected effects on decision satisfaction. People usually prefer to make their own decisions rather than having others decide for them (e.g., Stotland & Blumenthal, 1964; Perlmuter & Monty, 1977; Taylor, Lichtman, & Wood, 1984; Gilovich & Medvec, 1995; Botti, Orfali, & Iyengar, 2009). However, this preference is seemingly eliminated when people are forced to choose among undesirable alternatives. Choosing among negative options lowers confidence and satisfaction with the outcome (Burger, 1989; Beattie, Baron, Hershey, & Spranca, 1994; Dhar, 1997), often so much so that people actively avoid making a decision at all (Botti & Iyengar, 2006). For example, research suggests that, when all of the alternatives are undesirable (e.g., two bad meal options), people are more satisfied with the outcome when someone else chooses for them than when they make the choice themselves (Botti & Iyengar, 2004). This disutility of choosing for oneself comes from the unpleasant process of focusing on the disadvantages of each outcome (see also Botti & McGill, 2006). Negative options, it seems, eliminate the benefits of choice.

Notably, all of the previous work in this area focused on decisions framed as choices – participants in these studies were asked to *choose* the option they most preferred. This brings us to the second major variable we consider: the framing of the decision. The person deciding between a chicken or steak entrée can see their decision as a choice (“I choose the chicken”) or

as a rejection ("I reject the steak"). The exact same options, with the exact same outcome, might be experienced differently when framed as a choice rather than as a rejection.

Therefore, it may not always be the case that people are less confident when deciding among negative stimuli; rather, this may only occur when the decision is construed as a choice rather than as a rejection. We posit that the unpleasantness of deciding between negative options is partially derived from the mismatch with the valence of the decision frame. People will feel more confident in their decision if the frame can be changed by matching a salient attribute (valence) of the decision frame with that of the options: *reject* the *undesired* option.

Although the frames are necessarily identical in terms of outcome for binary choices, that does not mean that people think of them identically (Hubert, Neale, & Northcraft, 1987; Park, Jun, & MacInnis, 2000). Positive frames (i.e., "choose") highlight positive attributes, whereas negative frames (i.e., "reject") highlight negative attributes (Houston, Sherman, & Baker, 1991; Shafir, 1993). Choosing may bring about more intuitive thinking, whereas rejecting more deliberative thinking (Sokolova & Krishna, 2016; Nagpal & Krishnamurthy, 2008). More importantly, recall that previous research had suggested that a focus on negative attributes diminishes the utility of choosing between negative options (Botti & Iyengar, 2004). That is only part of the story. We propose that the match between the decision frame and the choice context (e.g., positive vs. negative options), what we will call *attribute matching*, determines the psychological response to the decision itself.

To our knowledge, only two published studies have come close to testing this hypothesis. Meloy and Russo (2004, Studies 2a and 2b) asked participants with to either choose or reject between positive options (e.g., good employees and good courses) or negative options (e.g., bad employees and bad courses), and then measured decision confidence. Their results showed

mixed support for our hypotheses: Although people were more confident when choosing between two positive options than when rejecting between them (as we would predict), they were equally confident when choosing between two negative options as when rejecting between them (as we would not predict). We believe these results, which were not of primary concerns to the authors, did not properly test our hypothesis in part because the authors did not cleanly manipulate valence, as, for example, the good employee had some negative features (e.g., called a "plodder") and the bad employee had some positive ones (e.g., said to have periods of "above average productivity"). In our studies, we avoid these problems by more cleanly manipulating valence, and demonstrate the robustness of our findings across multiple stimuli and multiple studies with large samples. More importantly, we propose and present evidence for a mechanism for this pattern of findings, generalize beyond simply valence matching and subjective decisions, and investigate not only reported confidence, but also perceived consensus in decision making.

We focus on a decision-making context in part because it is so broadly applicable. If a decision is experienced more positively, it should spillover into at least two critical domains: decision confidence and consensus estimates. Regarding choice confidence, decision makers are necessarily trying to identify the correct answer, but any decision will come with some sense of uncertainty. The experience of confidence in a decision can guide how sensitive people are to other constraining information and thus their subsequent behavior (e.g., Simmons & Nelson, 2006). Accordingly, it would be both important (and surprising) if merely increasing the apparent match between options and frame could operate on confidence.

Regarding consensus estimation, this construct has been at the core of advances in the understanding of social judgment. Starting with initial work on the false consensus effect (Ross, Greene, & House, 1977) there has been recognition that people first look inward when asked to

make judgments about others. This projection tendency, whether rational or irrational (Dawes, 1990; Krueger & Clement, 1994), springs from egocentrism. The tendency is consequential. For example, people misjudge the thirst and hunger of others depending on their own state (Van Boven & Loewenstein, 2003) and misjudge the humor of a new joke based on their own prior exposure (Campbell et al., 2014). Those misjudgments spring from transient personal states, unaccounted for when characterizing others in a different state. Again, notably, and most interestingly, we think that this might occur as a result of merely manipulating the decision frame. People might experience the transient confidence from the match of valence and frame, and infer that their decision will be more popular than it actually is.

Our studies investigate how this attribute matching can influence confidence and consensus estimates. In Study 1, we establish evidence for the attribute matching hypothesis, showing that people are more confident in, and believe that others are more likely to agree with, choices between positive options and rejections between negative options than choices between negative options and rejections between positive options. Study 2 generalizes this effect to a new domain, while showing that the effect is stronger when the choice options are extreme rather than moderate. In Study 3, we examine whether the effects arise because the speed and ease of matched decisions inspire greater confidence. In Study 4, we show that the effect hinges on a lack of awareness, as it is eliminated when people are warned that attribute matching might influence their confidence. Finally, Study 5 explores whether the matching hypothesis extends to objective judgments (e.g., calorie estimations). We conclude with a discussion of alternative explanations and future work for attribute matching.

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the respective study descriptions; we analyzed our data only

after collection had finished (Simmons, Nelson, & Simonsohn, 2010). Every study in this paper was pre-registered with the Open Science Framework. Pre-registered hypotheses, sample sizes, materials, procedures, exclusion criteria, and analysis plans, and full data sets for each study are linked within each study's description.

Study 1

Each of the studies manipulating decision frame use a similar paradigm. In this study, participants were asked to make a decision between two similarly likable options. For some people this decision was expressed as a choice, whereas for others it was framed as a rejection. Additionally, we varied the valence of the pairs of options. Some pairs consisted of two desirable options and some pairs consisted of two undesirable options. In this way we orthogonally manipulated the frame of the decision and the valence of the targets, creating matched and mismatched pairings. Moreover, our mixed design allowed all participants to experience both types of pairings. Choosing from attractive options and rejecting from unattractive options were the matched-valence trials. All participants made their decision and then reported their decision confidence and their estimate of the percentage of others who would have made the same decision. (Pre-registration and data: https://osf.io/tc7h4/?view_only=b1b5ff4d92bc47fda98b11ae3542f1c9 .)

Method

Participants. We recruited 2,519 participants from a private research company (M_{age} = 48.5, 51.5% female) to complete a survey about preferences. We determined this sample size to be all participants in one survey session. This session consisted of multiple unrelated studies strung together, and ran until at least 1,500 participants had passed the attention check in the first study of the set. We decided in advance to analyze only the data of those who passed the check.

We did not analyze the data for Study 1 until that threshold had been met and data collection had ceased[1].

Materials, procedure, and design. Participants were asked to imagine they were selecting models for an upcoming advertising campaign. They viewed pairs of women's headshots from Pochon, Riis, Sanfey, Nystrom, and Cohen (2008), which were pretested to be attractive or unattractive. The pairs were designed such that attractiveness did not significantly differ within each pair (see Pochon et al., 2008) for more information about the pretest). Specifically, participants saw 16 pairs of women's headshots: eight pairs of attractive women and eight pairs of unattractive women, presented in a randomized order. For approximately half of participants the decision was framed as a choice (i.e., "Which woman would you choose?") and for the remaining participants the decision was framed as a rejection (i.e., "Which woman would you reject?"). See Figure 1 for a sample attractive pair in the choose condition.

After each selection, participants were reminded of their answer and reported how confident they were in their decision (1 = not at all confident, 9 = extremely confident) and, as a measure of perceived consensus, what percentage of other people would make the same decision (sliding scale: 0% - 100%, with the marker starting at 50%). At the conclusion of the 16 trials, participants were given a brief attention check (Oppenheimer, Mevis, & Davidenko, 2009) and provided their age and gender.

Figure 1. A choice screen from the Choose/Attractive condition of Study 1.

Figure 1. A choice screen from the Choose/Attractive condition of Study 1.

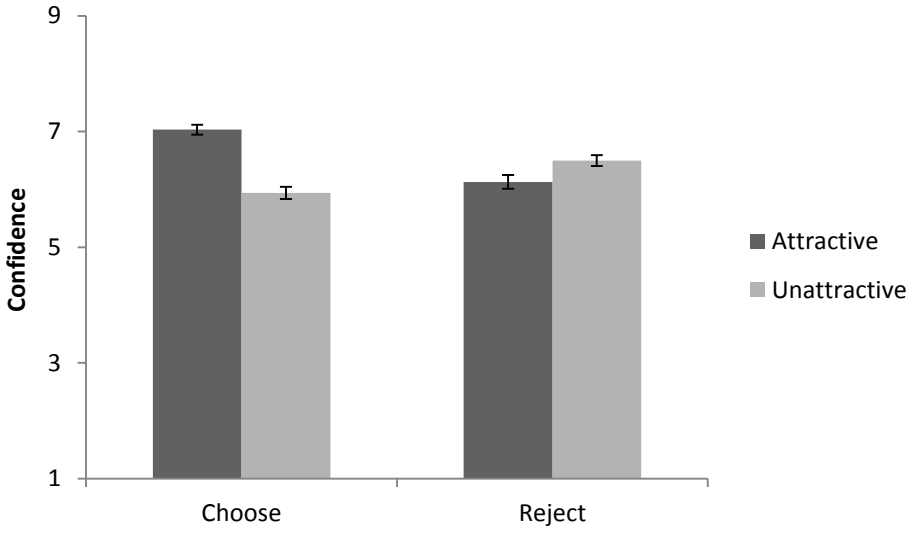
Results

In this study, 1,018 participants (40.4%) failed the attention check and were excluded from analyses. Although this number may seem high, it is not atypical for this more naïve participant pool and the exclusion rule adheres to our preregistration plan.[2] Figure 2 plots the results.

Confidence. With the remaining 1,501 participants, we ran an ordinal least squares (OLS) regression, clustering standard errors at the participant level, with *frame*, for the valence of the decision frame (0.5 = choose, -0.5 = reject), and *option*, for the valence of the decision's options (0.5 = attractive faces, -0.5 = unattractive faces), as well as their interaction. A main effect of *option* emerged (b = 0.43, SE = 0.03 $p < .001$), as well as a main effect of *frame* (b = 0.16, SE = 0.06, $p < .001$). Critically, these effects were qualified by our predicted interaction (b = 1.10, SE = 0.06, $p < .001$): for attractive targets, participants who were choosing the better

model were more confident in their decision (M = 7.04) than were the participants rejecting the worse model (M = 6.33), $t(1499) = 11.09$, $p < .001$. For unattractive targets, the effect reversed ($M_{choose} = 6.05$ vs. $M_{reject} = 6.45$), $t(1499) = 5.95$, $p < .001$).

Consensus. We then fit the same model for consensus estimates. As before, although *frame* ($b = 1.96$, SE = 0.55 $p < .001$) and *option* ($b = 0.73$, SE = 0.29, $p = .013$) were both significant predictors of consensus estimates, these effects were qualified by the predicted crossover interaction ($b = 9.19$, SE = 0.58 $p < .001$): for attractive trials, participants choosing the better model gave higher consensus estimates (M = 68.4%) than participants rejecting the worse model (M = 61.9%), $t(1499) = 10.19$, $p < .001$). However, this effect reversed for unattractive trials ($M_{choose} = 63.1\%$, $M_{reject} = 65.7\%$; $t(1499) = 4.42$, $p < .001$). In addition, when we entered confidence measures into the model, the interaction effect for consensus estimates was reduced (from $b = 9.19$ to $b = 2.35$), consistent with mediation, $z = 4.92$, $p < .001$.



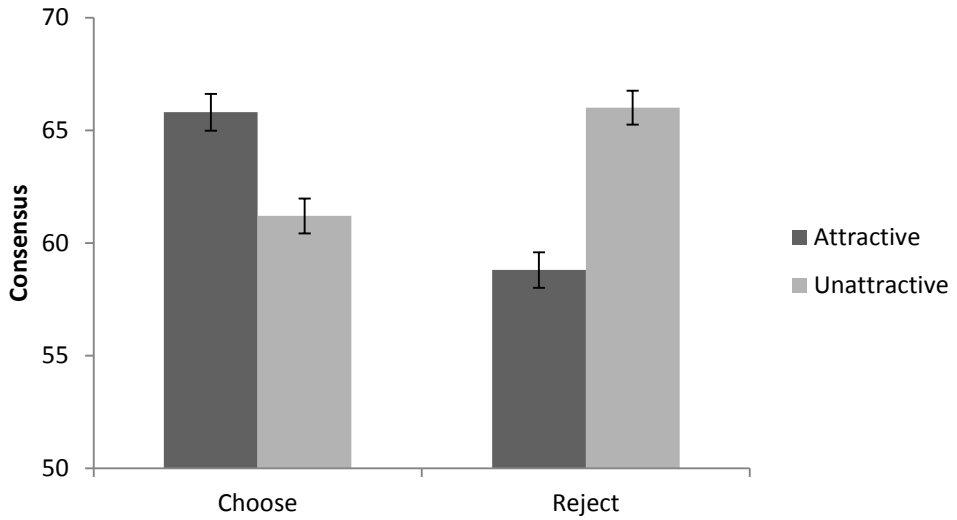


Figure 2. Matched trials (outer bars) showed higher decision confidence and perceived consensus than mismatched trials (inner bars) in Study 1.

Replication

To verify this initial demonstration of attribute matching was replicable, we recruited 303 participants from Amazon Mechanical Turk (M_{age} = 35.1, 65.6% female), with three participants (1.0%) failing the attention check and being excluded from analyses (Pre-registration and data: https://osf.io/bm6jr/?view_only=ff70d8b1f9da4c5cab8f16546e0340dc). We fit the same model used in Study 1 for confidence and consensus estimates and found an identical pattern: confidence and consensus estimates were significantly higher when the valence of the decision frame and the options matched than when they mismatched ($b_{confidence}$ = 1.46, SE = .06; $b_{consensus}$ = 11.19, SE = 1.18, *p*s < .001). As before, when we entered confidence measures into the model, the interaction effect for consensus estimates was reduced (from b = 11.19 to b = 3.71), again consistent with mediation, z = 3.86, p < .001. The full data set is posted along with the pre-registration above.

Discussion

When people were “choosing” between two attractive faces they were more confident in their decision and thought more people would agree with them than when they were “rejecting” between the exact same pair. This demonstrates a strong attribute matching effect: When people were asked to choose between options, we found that people were more confident in their decisions for attractive pairs than for unattractive pairs. However, this finding reversed under a reject frame. These matching effects on consensus estimates were then partially explained by the matching effects on confidence. Notably, the same results were obtained in a replication with a different population.

One concern might be in the extremity of the stimuli. To avoid the null and inconsistent results of Meloy & Russo (2004), our stimuli were specifically chosen to be at the ends of the attractiveness scale; however, perhaps in doing so, we unfairly stacked the deck in favor of the effect. Would attribute matching still emerge if the attribute, while still salient, were not so extreme? We carry out this more conservative test in Study 2.

Study 2

As in Study 1, we varied the valence of the decision frame (choose vs. reject) and the valence of the options (positive vs. negative). However, to test whether attribute matching would emerge with less extreme stimuli, we included four levels of valence: extremely positive, slightly positive, slightly negative, and extremely negative (see Table 1 for exact stimuli). We predicted that attribute matching would still emerge for the less extreme stimuli, albeit in smaller magnitudes. (Pre-registration and data: https://osf.io/expw5/?view_only=5e83d84d83b443198cf808f9d86fde8b.)

Method

Participants. We decided in advance to recruit 300 participants from Amazon Mechanical Turk, and received 301 (M_{age} = 36.1, 55.0% female).

Materials, procedure, design. Participants viewed 20 pairs of words in a randomized order. The words were chosen based on valence ratings provided by Bellezza, Greenwald, and Banaji (1986) and paired such that valence did not significantly differ within each pair. The 20 trials consisted of five pairs of extremely positive, slightly positive, slightly negative, and extremely negative words each (see Table 1). As in Study 1, half of participants were asked to indicate their preference by choosing the word they preferred, and half were asked to do so by rejecting the word they did not prefer. (Readers interested in the choice share within each pair are referred to the Supplementary Materials.) After making each choice, participants completed the same confidence and consensus measures from Study 1. At the conclusion of the study, participants received the same attention check as in Study 1 and provided their age and gender.

Very Negative**	Slightly Negative	Slightly Positive	Very Positive**
murderer vs. tumor	thorn vs. jealousy	circus vs. world	joy vs. kiss
poison vs. slaughter	snob vs. beggar	fur vs. privacy	pleasure vs. vacation
war vs. maggot	useless vs. wasp	knowledge vs. learn	family vs. laughter
cancer vs. funeral	rage vs. stress	water vs. employment	paradise vs. sunrise
lice vs. suicide	putrid vs. stupid	earth vs. improve	romantic vs. love

Table 1. Word pairs used in Studies 2-4 (**indicates use in Studies 3-4).

Results and Discussion

Fourteen participants (4.7%) failed the attention check and were excluded from analyses. Figure 3 plots the results.

Confidence. With the remaining participants, as in Study 1, we ran an OLS regression, clustering standard errors at the participant level, with *frame*, for the valence of the decision frame (0.5 = choose, -0.5 = reject), and *option*, for the valence of the decision's options (0.5 = attractive faces, -0.5 = unattractive faces), as well as their interaction. Although *frame* (*b* = 0.17,

SE = 0.12, p = .147) and *option* (b = -0.07, SE = 0.06, p = .249) were not significant predictors of confidence, our predicted crossover interaction was highly significant (b = 2.10, SE = 0.12, p < .001): on positive trials, participants choosing the better word were more confident (M = 6.92) than participants rejecting the worse word reported greater confidence in their choice (M = 6.00), $t(285) = 6.63$, $p < .001$, an effect that reversed for negative trials ($M_{choose} = 5.79$, $M_{reject} = 7.02$; $t(285) = 9.09$, $p < .001$). Supporting the idea that attribute extremity moderates the matching effect, the three-way interaction among frame, option, and extremity was highly significant, b = 1.43, SE = 0.17, p < .001. Accordingly, attribute matching occurred even for the slightly positive and slightly negative word pairs, but was less pronounced, $b_{interaction} = 1.38$, SE = 0.13, p < .001.

Stated simply, as in Study 1, choosing between positive options led to higher confidence that rejecting between the same options; choosing between negative items, led to lower confidence than rejecting between the same options. Additionally, Study 2 contributes the new insight that that reversal interaction is stronger for extreme option pairs than for mild option pairs, but the interaction remained significant for both.

Consensus. We then repeated this analysis with consensus estimates. The effect of *frame* was not significant (b = -0.53, SE = .99, p = .590), whereas the effect of *option* was (b = -2.01, SE = 0.56, p < .001). Critically, however, our predicted interaction was highly significant (b = 13.74, SE = 1.12, p < .001): on positive trials, participants choosing the better word gave higher consensus estimates (M = 65.0%) than participants rejecting the worse word (M = 57.6%), $t(285) = 6.61$, $p < .001$. However, this effect reversed for negative trials ($M_{choose} = 56.1\%$, $M_{reject} = 62.5\%$; $t(285) = 5.49$, $p < .001$). As with confidence estimates, this pattern emerged, though smaller as predicted, for the non-extreme stimuli alone ($b_{interaction} = 7.61$, SE = 1.30, p < .001), and the three-way interaction was significant: b = 12.26, SE = 1.75, p < .001. As in Study 1,

when we entered both the confidence and consensus measures into the model, the frame/option interaction effect for consensus estimates is reduced (to $b = 3.65$ from $b = 13.74$), consistent with mediation, $z = 3.72$, $p < .001$.

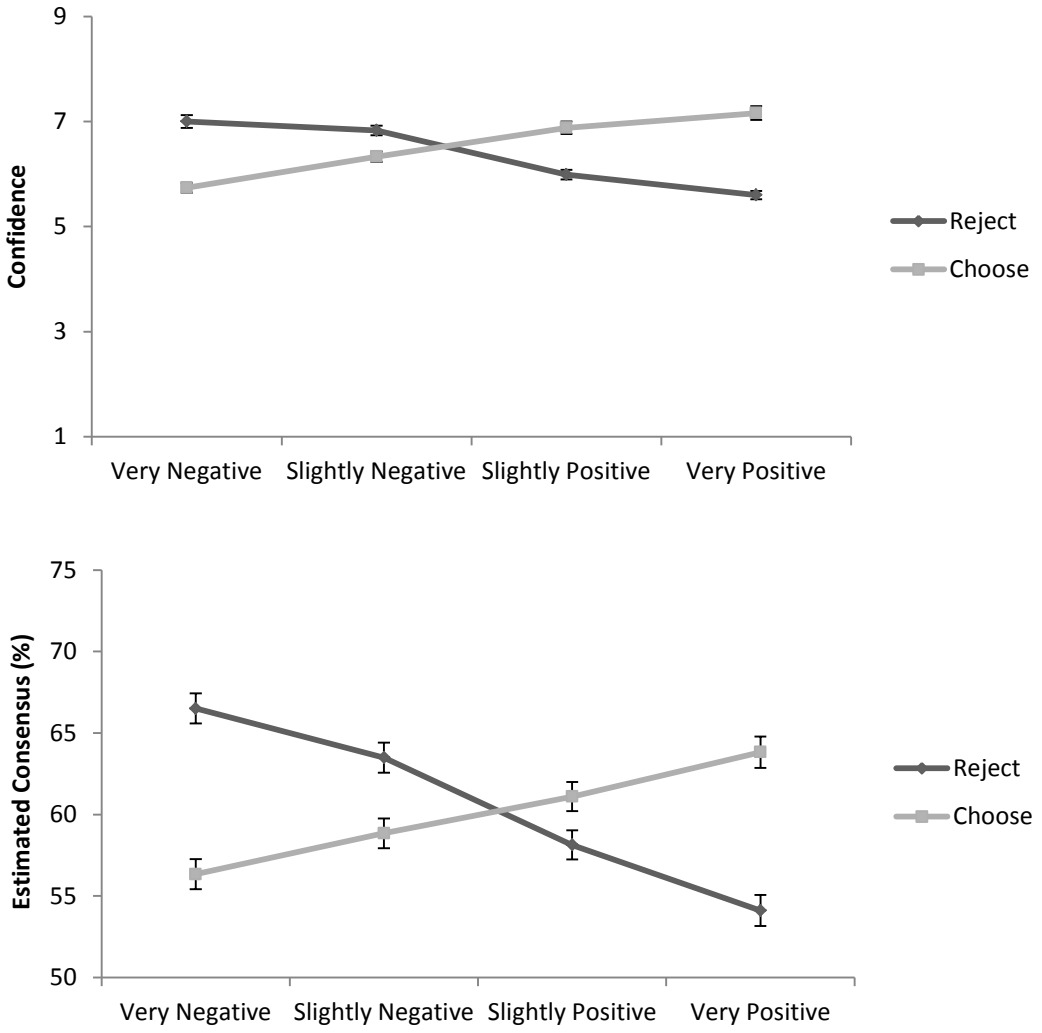


Figure 3. The size of the attribute matching effect varies with attribute intensity in Study 2.

Study 2 shows that attribute matching occurs not just when the attribute is at its extremes; even the only slightly positive and slightly negative word pairs significantly showed attribute matching, with the size of the effect varying with attribute intensity. In the next two studies, we turn to a possible mechanism for attribute matching: decision ease.

Study 3

Study 3 mirrored Study 2 except for two key differences. First, Study 3 only used the extremely positive and negative words, and second, it included both direct (Likert-like scale) and indirect (decision time) measures of decision ease Decisions on matching trials should feel easier to make, and be made faster, than those for mismatched trials. Furthermore, we predicted the standard matching effect from Study 2 would replicate, but, more importantly, that it would be fully mediated by our direct measure of decision ease. (Because of the unreliability of response time data [Fazio, 1990; Evans, Dillon, & Rand, 2015], we pre-registered our response time variable as exploratory. Pre-registration and data: https://osf.io/jk3mu/?view_only=e9dbb53c47a843d19d7ca5652d8c30d3)

Method

Participants. We recruited 421 participants from Amazon Mechanical Turk (M_{age} = 31.1, 57.2% female). We had pre-determined to recruit 400 participants in total.

Materials, procedure, and design. Participants viewed the five extremely positive and five extremely negative word pairs from Study 2, in a randomized order. As in Study 2, participants were asked either to choose the word they prefer or reject the word they did not prefer. We surreptitiously recorded the amount of time participants spent on this decision page. Next, all participants were asked how easy making their decision felt (1 – Very difficult, 9 – Very easy), before also responding to the same confidence and consensus measures we used in the previous studies. After the last trial, participants received the same attention check as in previous studies and reported their age and gender.

Results and Discussion

Sixty (14.3%) participants failed the attention check and were excluded from analyses.

Ease. With the remaining participants, we fit the same model as in Study 2 first to the ease measure (*Frame*: -0.5 = choose, 0.5 = reject; *option*: -0.5 = negative, 0.5 = positive). Neither f*rame* (b = -0.06, SE = 0.09 p = .511) nor *option* (b = -0.06, SE = 0.07, p = .432) were significant predictors of ease; however, our predicted crossover interaction was (b = -2.84, SE = 0.14, p < .001): on positive trials, participants choosing the better word reported the decision was easier to make (M = 5.73) than participants rejecting the worse word (M = 4.25), t(359) = 13.22, p < .001, an effect that reversed for negative trials (M_{choose} = 4.37, M_{reject} = 5.73; t(359) = 11.67, p < .001).

Response Time. Per our pre-registration, due to the nature of response times, we first excluded any responses less than 200ms and then log-transformed the resulting data before conducting our analyses (Wheelan, 2008); however, for ease of understanding, we report the raw values in text.[3]We then fit the model for our transformed response time variable. Here, *frame* was a significant predictor of response time (b = 0.09, SE = 0.01, p = .046) and *option* was not (b = -0.01, SE = 0.01, p = .117). More importantly, however, our predicted interaction again obtained (b = 0.12, SE = 0.01, p < .001). In line with our hypothesis, on positive trials, participants who chose the words they preferred chose faster (M = 5.39s) than did participants who rejected the words they did not prefer (M = 7.46s), t(357) = 8.52, p < .001. However, there was no significant difference on negative trials (M_{choose} = 6.56s, M_{reject} = 7.62s), t(367) = 1.28, p = .203.

Confidence. We next fit the model for reported confidence. As with reported ease, neither *frame* (b = -0.14, SE = 0.11, p = .231) nor *option* (b = 0.11, SE = 0.08, p = .182) were significant predictors of confidence; however, our predicted interaction was (b = -2.61, SE = 0.17, p < .001): on positive trials, participants who chose the words they preferred reported higher confidence in making their choice (M = 7.50) than did participants who rejected the words

they did not prefer (M = 6.06), $t(359) = 10.50$, $p < .001$, an effect that reversed for negative trials ($M_{choose} = 6.08$, $M_{reject} = 7.25$; $t(359) = 8.15$, $p < .001$). When we entered ease into the model for confidence, the frame/option interaction was reduced (from $b = -2.61$ to $b = -0.18$), suggesting mediation, $z = 1.99$, $p = .047$.

Consensus. Finally, we repeated this analysis with consensus estimates, and found a similar pattern: *frame* was not a significant predictor of consensus estimates ($b = -1.08$, SE = 0.92, $p = 243$), *option* was ($b = -3.65$, SE = 1.67, $p < .001$), but, critically, our predicted interaction obtained ($b = -14.30$, SE = 1.34, $p < .001$): on positive trials, participants who chose the words they preferred gave higher consensus estimates in making their choice (M = 64.4%) than did participants who rejected the words they did not prefer (M = 56.1%), $t(359) = 7.41$, $p < .001$, an effect that reversed for negative trials ($M_{choose} = 60.9\%$, $M_{reject} = 66.9\%$; $t(359) = 5.16$, $p < .001$). When we entered measures of ease and confidence into the model for consensus, the interaction effect is reduced to non-significance (from $b = -14.30$ to $b = -2.27$), suggesting full mediation.

Study 3 presents preliminary evidence for decision ease as a mechanism behind attribute matching effects on confidence. In addition to replicating the attribute matching effect from Study 2, participants' reports of how easy making the decisions felt, as well as how quickly they made those decisions, fully mediated their reported confidence. Along these lines, we hypothesize that participants misattribute this change in decision ease to their confidence in their preferences, instead of to our manipulations. However, rather than affecting subjective ease, it could be that our manipulations affect how objectively easy the decisions were to make. To disentangle these two possibilities in Study 4, we prompt some participants to consider this irrelevant source of experiential information. If this is a matter of misattribution, this prompt

should attenuate the effect; if instead the effect emerges only because matched decisions are easier, then this prompting should have no effect.

Study 4

The purpose of Study 4 was to more directly test the hypothesis that the increase in confidence and consensus estimates on matched trials comes from a misattribution of decision ease from attribute matching, and not from the decision becoming objectively easier to make. Study 4 used the same materials and procedure as Study 3, but without the additional measure of ease. (Response time was not measured in this study.) Instead, we introduced a third factor: some participants were alerted that the valence of the frame and the options may have made the decision feel easier prior to their reports of confidence and consensus. We predicted that participants who did not receive this notice would show the attribute matching effect, whereas those who did would properly attribute their increase in ease to our manipulations and show no effects. (Pre-registration and data: https://osf.io/gdrvs/?view_only=6443e8fa7ae04ee1a9c86d6405ce9f4a.)

Method

Participants. We recruited 406 participants from Amazon Mechanical Turk (M_{age} = 32.9, 52.6% female). We had pre-determined to recruit 400 participants in total. We did not analyze the data until that threshold had been met and data collection ceased.

Materials, procedure, and design. As in Study 3, participants viewed five extremely positive and five extremely negative word pairs in a randomized order, and were randomly assigned either to choose the word from each pair that they prefer or to reject the word they do not prefer. As in previous studies, all participants reported their confidence and perceived consensus. However, after making their selection but before reporting their confidence,

consistent with the approach taken by Cesario, Grant, and Higgins (2004), half of participants read: "Before continuing, please consider the following: Past research suggests that phrasing a decision positively [negatively] could affect how easy your decision seems, *depending on* the positivity or negativity of the options." Participants in these alerted conditions saw the message on every trial, and could not proceed to the next page for two seconds. At the conclusion of the study, participants received the same attention check as in previous studies and provided their age and gender.

Results and Discussion

Sixty-five participants (16%) failed the attention check and were excluded from analyses. Figure 4 plots the results.

Confidence. First, for confidence, we fit the same model as in Study 3 but with *alert* (-0.5 = standard, 0.5 = alerted) as a third predictor and its two-way interactions with *option* and *frame* and the three-way interaction. *Frame* (b = -0.28, SE = 0.09, p = .002) and *option* (b = 0.34, SE = 0.06, p < .001) were both significant predictors of reported confidence. As in previous studies, these effects were qualified by the predicted crossover interaction (b = -0.94, SE = 0.12, p < .001), but, more importantly, the predicted three-way interaction was also significant, b = 2.76, SE = 0.23, p < .001. To probe the nature of this interaction further, we discuss the standard and alerted conditions separately.

In the standard conditions, without the alerting manipulation, attribute matching replicated as predicted: on positive trials, participants reported greater confidence when choosing (M = 7.38) than rejecting (M = 5.82), t(163) = 7.53, p < .001; whereas on negative trials, participants reported greater confidence when rejecting (M =7.10) than choosing (M = 6.49), t(163) = 3.23, p = .001. In the alerted conditions, however, the results were very different: for

positive trials participants reported similar levels of confidence whether they were choosing (M =7.24) or rejecting (M = 6.96), $t(174) = 1.84$, $p = .068$. On negative trials, an unpredicted significant difference did emerge, but in the direction opposite of what was observed in the control condition ($M_{choose} = 6.65$, $M_{reject} = 6.13$; $t(174) = 2.58$, $p = .011$).

Consensus. We then fit the same model for consensus estimates. Neither f*rame* ($b = -1.65$, SE = 0.87, $p = .060$) nor option valence ($b = 0.89$, SE = 0.61, $p = .146$) significantly predicted estimated consensus. As in previous studies, however, our predicted crossover interaction was highly significant ($b = -3.84$, SE = 0.61, $p = .002$), and more importantly, the predicted three-way interaction was also significant, $b = 22.93$, SE = 2.44, $p < .001$. To probe the nature of this interaction further, we again discuss the standard and alerted conditions separately.

In the standard conditions, without alerting, attribute matching replicated as predicted: on positive trials, participants estimated higher consensus when choosing (M = 65.5%) than rejecting (M = 54.7%), $t(162) = 5.91$, $p < .001$; whereas on negative trials, participants estimated higher consensus when rejecting (M =68.6%) than choosing (M = 63.6%), $t(163) = 2.99$, $p = .003$. In the alerted conditions, however, the results were very different: participants reported the same level of consensus on positive trials, regardless of whether they were choosing (M = 64.8%) or rejecting (M = 66.3%), $t(174) = .97$, $p = .336$. On negative trials, a significant difference did emerge, but, again, in the direction opposite that observed in the control conditions ($M_{choose} = 62.8\%$, $M_{reject} = 57.2\%$; $t(174) = 3.11$, $p = .002$).

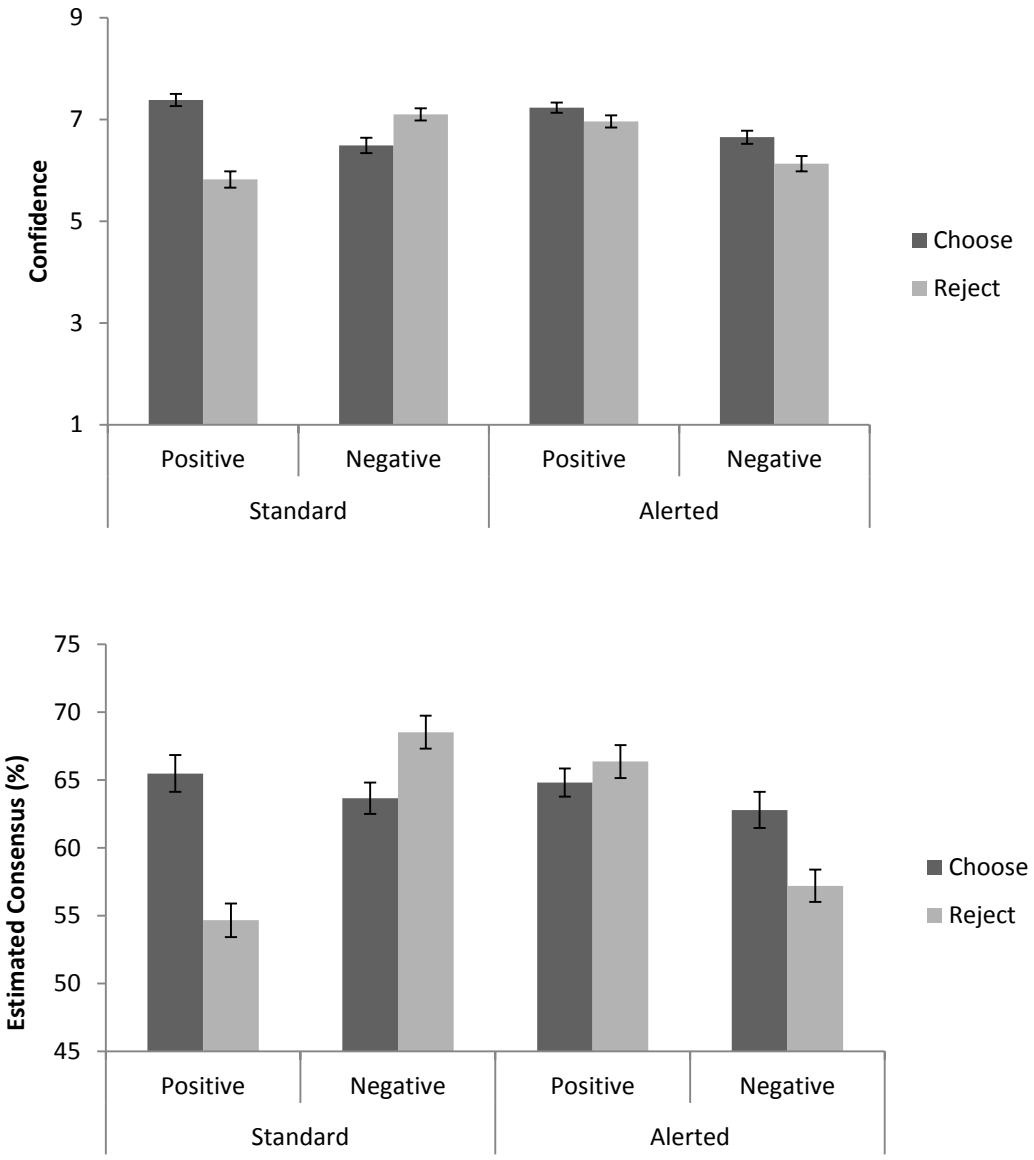


Figure 4. Attribute matching fails to emerge after alerting in Study 4.

Replication

Because of the theoretical importance of Study 4 to the present paper, we ran a direct replication with the exact same materials to verify its reliability (pre-registration and data: https://osf.io/4tzig/?view_only=082934b269e64ae49051b982f2ea26a4). We recruited 502 participants from Amazon Mechanical Turk; forty-four (8.8%) failed the attention check and were excluded from analyses. We fit the same model as before: the three-way interaction again

emerged for both confidence (b = 2.75, SE = 0.22, p < .001) and consensus (b = 22.40, SE = 2.31, p < .001). The standard conditions showed the usual frame/option interaction ($b_{confidence}$ = -2.30, SE = 0.17, p < .001; $b_{consensus}$ = -15.14, SE = 1.79, p < .001). However, the alerted conditions did not ($b_{confidence}$ = .45, SE = 0.14, p = .002; $b_{consensus}$ = 7.26, SE = 1.47, p < .001).

Discussion

Study 4 offers further support for decision ease as a mechanism behind attribute matching. When participants were given a reason to attribute their increased ease of decision making to the valence of the frame and options, the effect disappeared (or even reversed, in the replication). Importantly, that warning did not indicate the direction of the effect, suggesting that an altered attribution of ease, rather than simply demand, is driving the effect.

However, the results of Studies 1 - 4 speak only to the robustness of attribute matching within preferences: these subjective judgments are much more susceptible to changes in experiential information than objective judgments, with more declarative information to call upon (Schwarz, 1998). For a more conservative test of attribute matching exploring a different domain, Study 5 employs a judgment task that has correct answers in a domain familiar to participants.

Study 5

Study 5 utilized a similar design as Studies 1 and 2, but moved beyond valence as the attribute being matched, into a more objective domain. Instead of seeing positive and negative stimuli and being asked for their preferences, participants saw pairs of high- and low-calorie foods and were either asked which food has more calories or which food has fewer, a question with a known answer. Deciding which high-calorie food has more calories and which low-calorie

food has fewer were the matched trials. (Pre-registration and data: https://osf.io/c7xez/?view_only=debc8c5fbf6d499d9977075be3317668.)

High-Calorie Foods	Low-Calorie Foods
Double cheeseburger vs. Medium pepperoni pizza	Baby carrot vs. Celery stick
12 buffalo chicken wings vs. Small cheese pizza	4 oz. of light orange juice vs. 4 oz. of light lemonade

Table 2. Food pairs used in Study 5.

Method

Participants. We recruited 408 participants from Amazon Mechanical Turk (M_{age} = 34.7, 62% female). We had pre-determined to recruit 400 participants in total. We did not analyze the data until that threshold had been met and data collection ceased.

Materials, procedure, and design. Participants viewed two pairs of high-calorie foods and two pairs of low-calorie foods, in a randomized order. Foods were determined to be high- or low-calorie based on a pretest, in which 51 Amazon Mechanical Turk participants were asked to estimate the caloric content of each food separately. See Table 2 for the four food pairs. The pairs were then constructed such that the median estimates were approximately equal. For these four food pairs, study participants were randomly assigned to determine which food has more calories or which food has fewer calories. After making each selection, participants completed the same confidence and consensus measures from the previous studies. We also asked participants how easy the decision felt to make, as in Study 3, to ensure any effects found here have a similar mechanism. At the conclusion of the study, participants received a brief attention check ("Which food has more [fewer] calories? 3 grapes vs. Grilled cheese sandwich"), reported if they were vegetarian as well as if they were on a diet, and provided their age and gender.

Results and Discussion

Eleven participants (2.7%) failed the attention check and were excluded from analyses.

Ease. With the remaining participants, we ran an OLS regression, clustering standard errors at the participant level, with *frame*, for the valence of the decision frame (-0.5 = more, 0.5 = less), and *option*, for the valence of the decision's options (0.5 = high-calorie foods, -0.5 = low-calorie foods), as well as their interaction. We found no effect of *frame* (b = -0.07, SE = 0.10, p = .468), although a significant effect of *option* (b = 0.42, SE = 0.07, p < .001) did emerge. More importantly, our predicted interaction was also significant (b = 0.75, SE = 0.14, p < .001): on high-calorie trials, participants who were asked which food has more calories reported the decision was easier to make (M = 4.55) than did participants who were asked which has fewer (M = 4.11), $t(395) = 3.91$, p < .001. However, this effect reversed when participants were asked about low-calorie foods (M_{more} = 4.60, M_{fewer} = 4.91; $t(395) = 2.35$, p = .019).

Confidence. We next ran the same analysis for reported confidence. Again, we found no effect of *frame* (b = 0.07, SE = 0.14, p = .618), but did find a significant effect of *option* (b = 0.71, SE = 0.10, p < .001). More importantly, our predicted interaction was significant, as well (b = 0.90, SE = 0.19, p < .001): on high-calorie trials, participants who were asked which food has more calories gave higher confidence estimates (M = 5.39) than did participants who were asked which has fewer (M = 5.01), $t(395) = 2.37$, p = .018. However, this effect reversed for low-calorie trials (M_{more} = 5.65, M_{fewer} = 6.17; $t(395) = 3.00$, p = .003). In addition, when we added decision ease to the model, the interaction effect for reported confidence was reduced to non-significance (from b = 0.90 to b = 0.12), consistent with mediation, $z = 5.19$, p < .001

Consensus. Finally, we fit the same model for consensus estimates. Again, we found no effect of *frame* (b = -0.39, SE = 1.05, p = .713), but did find a significant effect of *option* (b = 5.85, SE = .86, p < .001). More importantly, the predicted crossover interaction was again highly significant (b = 7.66, SE = 1.73, p < .001): participants who were asked which food had more

calories gave higher consensus estimates for choices made for high-calorie food pairs (M = 62.0%) than for low-calorie pairs (M = 57.8%), $t(395) = 3.15$, $p = .002$). However, this effect reversed for participants who were asked which foods had fewer calories ($M_{more} = 64.1\%$, $M_{fewer} = 67.5\%$; $t(395) = 2.47$, $p = .014$). In addition, when we entered reported confidence into the model, the interaction effect for consensus estimates was reduced (from $b = 7.66$ to $b = 3.63$), consistent with mediation, $z = 5.07$, $p < .001$.

Study 5 shows that the effects of attribute matching are not limited to judgments of subjective preference, but emerge even when people are judging stimuli on objective dimensions. To our knowledge, this is the first demonstration of a matching effect on objective judgments with correct answers. Even when asked to make objective judgments of caloric content, participants' confidence and perceived consensus were higher when the decision frame and the food pairs matched on their salient attribute.

General Discussion

No one wants to face undesirable options, but the present paper demonstrates that people might still feel good about their decisions amongst them. We propose that this can happen through attribute matching: When a salient attribute of the decision frame matches a salient attribute of the options, the decision feels easier to make, which increases reported confidence and perceived consensus. In Study 1, we showed that participants were more confident in their preference and perceived greater consensus around it when they were choosing from attractive faces or rejecting from unattractive faces than when the frame and the options did not match in valence. Study 2 found that this matching effect emerged even with only slightly valenced options. Study 3 directly tested a possible mechanism by measuring decision ease and response time and finding strong evidence of attribute matching and mediation of confidence estimates;

hence, in Study 4, notifying participants that the valence of the frame and options may affect their decision ease appeared to correct their misattribution of that ease and eliminate the effect. Finally, in Study 5, we extended this finding to objective judgments. Despite experiential information now having to compete with much more declarative information, we still found evidence of attribute matching with calorie judgments in Study 5.

One could, however, propose a few alternative explanations, instead of decision ease. For example, perhaps participants in mismatched trials were answering a different question than the one asked (i.e., response substitution, [Gal & Rucker, 2011]). When shown a pair of faces participants want to answer "how attractive are these faces?" but are instead asked only "how confident are you in your decision?" They substitute their response to the former when answering the latter. We assessed this account in a modified version of Study 1: we selected the two top-performing attractive and unattractive trials from the choose conditions of Study 1 (to conserve statistical power; N = 401) and varied whether participants first had the opportunity to rate how well-suited either face was for the advertising campaign before rating how easy the decision felt to make and their confidence in their decision (pre-registration and data: https://osf.io/zg6b5/?view_only=69529f4e3b4a4323bc4b757ac5f17e0). Contrary to a response substitution account, we again found evidence of attribute matching, regardless of whether participants had the chance to express their attitudes towards the stimuli first (see the Supplementary Materials for a full description of the methods and results, as well as a file drawer of three additional studies not included in the present paper).

A second alternative explanation often put forth for matching effects is positive mood, rather than decision ease (e.g., Cesario, et al., 2004). Although we did not directly measure mood in our studies, most of our studies included explicitly negative stimuli (e.g., the words "murder"

and “maggot”), which is not conducive to positive mood. Accordingly, this seems like a relatively unlikely account for the findings.

A broader step for further investigating could look into the connection of some otherwise disparate findings in the literature that relate to our findings. For example, work on regulatory fit (Higgins, 2000) claims that “when [people] use goal pursuit means that fit their regulatory orientation” (p. 1219) this fit generates its own utility (Higgins, 2003), increasing decision confidence (Cesario, Grant, & Higgins, 2004) and satisfaction (Idson, Liberman, & Higgins, 2000). Construal level theory (Trope & Liberman, 2000) research has shown when an option’s features (construals) match each other, decision makers view those options more positively (Todorov, Goren, & Trope, 2007; Fujita et al, 2006; Kim, Rao, & Lee, 2009). It could be that, in abstract, these papers and literatures are each their own form of attribute matching: each matching attributes from different parts of the decision process, with positive outcomes for the decision maker, due to increases in decision ease. The case of attribute matching we present here is a pure, robust, reliable form of what is at the core of these important literatures. Matching itself is understudied, yet ubiquitous, and ripe for future research.

References

Beattie, J., Baron, J., Hershey, J. C., & Spranca, M. D. (1994). Psychological determinants of decision attitude. *Journal of Behavioral Decision Making*, *7*, 129–144.

Bellezza, F. S., Greenwald, A. G., & Banaji, M. R. (1986). Words high and low in pleasantness as rated by male and female college students. *Behavior Research Methods, Instruments, & Computers*, *18*, 299-303.

Botti, S., & Iyengar, S. S. (2004). The psychological pleasure and pain of choosing: when people prefer choosing at the cost of subsequent outcome satisfaction. *Journal of Personality and Social Psychology*, *87*, 312.

Botti, S., & Iyengar, S. S. (2006). The dark side of choice: When choice impairs social welfare. *Journal of Public Policy & Marketing*, *25*, 24-38.

Botti, S., & McGill, A. L. (2006). When choosing is not deciding: The effect of perceived responsibility on satisfaction. *Journal of Consumer Research*, *33*, 211-219.

Botti, S., Orfali, K., & Iyengar, S. S. (2009). Tragic choices: Autonomy and emotional responses to medical decisions. *Journal of Consumer Research*, *36*, 337-352.

Burger, J. (1989). Negative reactions to increases in perceived personal control. *Journal of Personality and Social Psychology*, *56*, 246–256.

Burson, K. A., Larrick, R. P., & Lynch, J. G. (2009). Six of one, half dozen of the other expanding and contracting numerical dimensions produces preference reversals. *Psychological Science*, *20*, 1074-1078.

Campbell, T., O'Brien, E., Van Boven, L., Schwarz, N., & Ubel, P. (2014). Too much experience: a desensitization bias in emotional perspective taking. *Journal of Personality and Social Psychology*, *106*, 272.

Cesario, J., Grant, H., & Higgins, E. T. (2004). Regulatory fit and persuasion: Transfer from "feeling right." *Journal of Personality and Social Psychology, 86*, 388.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*, 1015.

Dawes, R.M. (1990). The potential nonfalsity of the false consensus effect. RM Hogarth (Ed.), Insights in Decision Making: A Tribute to Hillel J. Einhorn, University of Chicago Press, Chicago.

Dhar, R. (1997). Consumer preference for a no-choice option. *Journal of Consumer Research, 24*, 215–231.

Evans, A. M., Dillon, K. D., Rand, D.G. (2015). Fast but not intuitive, slow but not reflective: Decision conflict drives reaction times in social dilemmas. *Journal of Experimental Psychology: General, 144*, 951-966.

Fujita, K., Henderson, M., Eng, J., Trope, Y., & Liberman, N. (2006). Spatial distance and mental construal of social events. *Psychological Science, 17*, 278–282.

Gal, D., & Rucker, D. D. (2011). Answering the unasked question: Response substitution in consumer surveys. *Journal of Marketing Research*, *48*, 185-195.

Gilovich, T., & Medvec, V. H. (1995). The experience of regret: what, when, and why. *Psychological Review*, *102*, 379.

Higgins, E. T. (2000). Making a good decision: value from fit. *American Psychologist, 55*, 1217.

Higgins, E. T. (2005). Value from regulatory fit. *Current Directions in Psychological Science, 14*, 209-213.

Houston, D. A., Sherman, S. J., & Baker, S. M. (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.

Idson, L. C., Liberman, N., & Higgins, E. T. (2000). Distinguishing gains from nonlosses and losses from nongains: A regulatory focus perspective on hedonic intensity. *Journal of Experimental Social Psychology, 36*, 252-274.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American psychologist*, *39*, 341.

Kim, H., Rao, A. R., & Lee, A. Y. (2009). It's time to vote: The effect of matching message orientation and temporal frame on political persuasion. *Journal of Consumer Research, 35*, 877-889.

Krueger, J., & Clement, R. W. (1994). The truly false consensus effect: an ineradicable and egocentric bias in social perception. *Journal of Personality and Social Psychology*, *67*, 596.

Meloy, M. G., & Russo, J. E. (2004). Binary choice under instructions to select versus reject. *Organizational Behavior and Human Decision Processes*, *93*, 114-128.

Nagpal, A., & Krishnamurthy, P. (2008). Attribute conflict in consumer decision making: The role of task compatibility. *Journal of Consumer Research*, *34*(5), 696-705.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*, 867-872.

Park, C. W., Jun, S. Y., & MacInnis, D. J. (2000). Choosing what I want versus rejecting what I do not want: An application of decision framing to product option choice decisions. *Journal of Marketing Research, 37,* 187-202.

Perlmuter, L. C., & Monty, R. A. (1977). The importance of perceived control: Fact or fantasy? Experiments with both humans and animals indicate that the mere illusion of control significantly improves performance in a variety of situations. *American Scientist,* 759-765.

Petty, R. E., & Krosnick, J. A. (1995). Attitude strength: Antecedents and consequences. Mahwah, NJ: Erlbaum.

Pochon, J. B., Riis, J., Sanfey, A. G., Nystrom, L. E., & Cohen, J. D. (2008). Functional imaging of decision conflict. *The Journal of Neuroscience, 28,* 3468-3473.

Read, D., Frederick, S., Orsel, B., & Rahman, J. (2005). Four score and seven years from now: The date/delay effect in temporal discounting. *Management Science, 51,* 1326-1335.

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13*, 279-301.

Russo, J. E., Meloy, M. G., & Wilks, T. J. (2000). Predecisional distortion of information by auditors and salespersons. *Management Science, 46,* 13-27.

Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory and Cognition, 21*, 546-556.

Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition, 49,* 11-36.

Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2012). A 21 word solution. *Dialogue: The Official Newsletter of the Society for Personality and Social Psychology, 26*, 4-7.

Sokolova, T., & Krishna, A. (in press). Take It or Leave It: How Choosing versus Rejecting Alternatives Affects Information Processing. *Journal of Consumer Research*.

Stotland, E., & Blumenthal, A. L. (1964). The reduction of anxiety as a result of the expectation of making a choice. *Canadian Journal of Psychology*, *18*, 139.

Taylor, S. E., Lichtman, R. R., & Wood, J. V. (1984). Attributions, beliefs about control, and adjustment to breast cancer. *Journal of Personality and Social Psychology*, *46*, 489.

Thaler, R. H., & Sunstein, C. R. (2008). Nudge: Improving Decisions about Health, Wealth, and Happiness. New Haven, CT: Yale University Press.

Todorov, A., Goren, A., & Trope, Y. (2007). Probability as a psychological distance: Construal and preferences. *Journal of Experimental Social Psychology*, *43*, 473-482.

Trope, Y., & Liberman, N. (2000). Temporal construal and time-dependent changes in preference. *Journal of Personality and Social Psychology*, *79*, 876.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453-458.

Tversky, A., & Shafir, E. (1992). The disjunction effect in choice under uncertainty. *Psychological science*, *3*, 305-309.

Van Boven, L., & Loewenstein, G. (2003). Social projection of transient drive states. *Personality and Social Psychology Bulletin*, *29*, 1159-1168.

Footnotes.

1. The sample size of 1,500 was set for the purposes of a larger project that specifies that target for all contributing experiments. Notably, as part of that project, Study 1 was replicated in three other labs sampling from three other populations. Although the project's protocol precludes us from sharing those results at this time, we can say that the effects were highly significant and similar in magnitude to the reported study.

2. Although we pre-registered to only analyze the data from those who passed the attention check, in each study the effect was very similar in size and highly significant when analyzing all participants. All data are available at the OSF pages.

3. Using untransformed response times yields equivalent effect sizes and significance levels. We report transformed results in keeping with convention and our pre-registration.

Supplementary Materials

First, we present the decisions made by participants in Study 2, split by condition, to show our manipulation did not significantly influence what participants selected. Next, we present the results of four studies that were not included in the final paper: The first two were successful, but were excluded out of length concerns. The second two were not successful, but may still be informative.

1. Trial-Level Decisions from Study 2

For ease of comparison to those of the choose condition, the reject conditions' selections have been reversed to show their de facto choice. Note that on only one trial (trial nine) of twenty did the majority preference differ between the two conditions (likely due to being so close to 50%).

Trial Number (valence)	Frame	% Choosing A	% Choosing B
1 (extremely negative)	Choose	41.3	58.7
	Reject	42.4	57.6
2 (extremely negative)	Choose	68.0	32.0
	Reject	85.4	14.6
3 (extremely negative)	Choose	62.7	37.3
	Reject	64.9	35.1
4 (extremely negative)	Choose	38.7	61.3
	Reject	35.1	64.9
5 (extremely negative)	Choose	70.7	29.3
	Reject	70.2	29.8
6 (slightly negative)	Choose	72.7	27.3
	Reject	66.2	33.8
7 (slightly negative)	Choose	47.3	53.7
	Reject	28.5	71.5
8 (slightly negative)	Choose	42.0	58.0
	Reject	41.1	58.9
9 (slightly negative)	Choose	50.7	49.3
	Reject	41.1	58.9
10 (slightly negative)	Choose	33.3	66.7
	Reject	23.2	76.8
11 (slightly positive)	Choose	36.7	63.3
	Reject	22.5	77.5

12 (slightly positive)	Choose Reject	44.0 22.5	56.0 77.5
13 (slightly positive)	Choose Reject	76.7 69.5	23.3 30.5
14 (slightly positive)	Choose Reject	64.7 82.8	35.3 17.2
15 (slightly positive)	Choose Reject	56.0 61.6	44.0 38.4
16 (extremely positive)	Choose Reject	63.3 66.2	36.7 33.8
17 (extremely positive)	Choose Reject	71.3 62.3	28.7 37.7
18 (extremely positive)	Choose Reject	48.7 39.1	51.3 60.9
19 (extremely positive)	Choose Reject	75.3 67.5	24.7 32.5
20 (extremely positive)	Choose Reject	66.0 84.8	34.0 15.2

2.1. Response Substitution Study

Method

Participants. We decided in advance to recruit 400 participants from Amazon Mechanical Turk and received 401 participants (M_{age} = 30.6, 52.6% female).

Materials, procedure, design. From Study 1's materials, we selected the two pairs of attractive faces and the two pairs of unattractive faces that showed the largest attribute matching effect. All participants saw these four pairs and, to conserve statistical power, all participants were asked which woman they would choose (i.e., we did not manipulate decision frame in this study). All participants reported how easy the choice was to make and how confident they were in their choice. However, for half of participants, after choosing a woman but before reporting decision ease, they were asked to rate the attractiveness of the women in the pair (1 = very unattractive, 7 = very attractive). After completing all four trials, participants reported their age

and gender, and completed a brief attention check. (Pre-registration and data: https://osf.io/zg6b5/?view_only=69529f4e3b4a4323bc4b75a7ac5f17e0 .)

Results.

Sixty-eight participants (17.2%) failed the attention check and were excluded from analyses.

Ease. With the remaining participants, as in Study 1, we first fit a hierarchical linear model for reported decision ease. We defined two Level-1 variables (*attract*: -0.5 = no attractiveness question, 0.5 = attractiveness question, and *option*: -0.5 = unattractive, 0.5 = attractive), and included their interaction. These variables were then nested within participant in a random-intercept model. Consistent with our hypothesis that our previous results were not driven by response substitution, we found only a significant effect of *option* (b = 1.13, SE = 0.10, $p < .001$); whether participants were able to express their opinions about the women's attractiveness beforehand had no effect on its own (b = -0.06, SE = 0.11, p = .577) nor did it moderate the *option* effect (b = 0.03, SE = 0.20, p = .866).

Confidence. We then repeated this analysis for reported decision confidence. Again, only *option* emerged as a significant predictor (b = 1.68, SE = 0.12, $p < .001$); allowing some participants to express their opinions beforehand had no effects on its own (b = .21, SE = .14, p = .153) or as a moderator (b = 0.22, SE = 0.24, p = .357).

2.2. Punishments Study

We also ran a study using purely negative stimuli. Although Study 2-4 also have negative stimuli (negative words), we wanted to verify that attribute matching would still hold for negative stimuli outside of the artificial domain of those studies (i.e., word preferences). In this

study, we presented participants with pairs of very aversive (e.g., cleaning public bathrooms) and less aversive (e.g., doing your laundry) tasks in either the "choose" or "reject" frame used in Studies 1-4. (Pre-registration and data: https://osf.io/bg4j6/?view_only=9a4409a84bc84b19a22b683da299d657)

Method

Participants. We recruited 1,152 participants from the same private research company, and as part of the same larger project, as Study 1. We determined this sample size to be half of all participants in a survey session that required 1,500 attention check passes in a separate survey. We did not analyze the data until that threshold had been met and data collection ceased.

Materials, procedure, design. Participants viewed three pairs of very unpleasant tasks and three pairs of slightly unpleasant tasks, in a randomized order. The tasks' unpleasantness was determined by a pretest, in which 52 Amazon Mechanical Turk participants were asked to rate each task's unpleasantness separately. For these six pairs, participants were randomly assigned to report which task they would rather do or which task they would rather not do. After making each selection, participants completed the same confidence and consensus measures from the previous studies. At the conclusion of the study, participants received a brief attention check and reported their age and gender.

Results

We first excluded 398 participants (34.5%) for failing the attention check.

Confidence. With the remaining participants, we first fit a hierarchical linear model for confidence estimates using clustered standard errors. We defined two Level-1 variables—*frame*, for the valence of the decision frame (-0.5 = choose, 0.5 = reject), and *option*, for the valence of the decision's options (0.5 = less unpleasant, -0.5 = more unpleasant)—and included their

interaction. These variables were nested within participant in a random-intercept model. A main effect of *frame* was not significant (b = -0.05, SE = 0.08, p = .533), although a main effect of *option* was (b = 0.40, SE = 0.05, p < .001). As predicted, however, our predicted interaction was highly significant (b = -.62, SE = 0.09, p < .001): for the less pleasant tasks, participants who were choosing were more confident in their decision (M = 7.93) than were the participants rejecting (M = 7.56), $t(752) = 11.09$, p < .001. For the more unpleasant tasks, the effect reversed (M_{choose} = 7.22 vs. M_{reject} = 7.48), $t(752) = 4.17$, p < .001).

Consensus. We then fit the same model for consensus estimates. As before, there was no main effect of *frame* (b = -0.99, SE = 087, p = .252), whereas there was for *option* (b = -1.33, SE = 0.33, p = .033). The predicted crossover interaction was only marginally significant (b = -2.41, SE = 1.25, p = .054): for less pleasant tasks, participants who were choosing the better task estimated higher consensus around their decision (M = 63.8%) than were the participants rejecting the worse task (M = 61.6%), $t(752) = 2.07$, p = .033. For the more unpleasant tasks, however, there was no difference (M_{choose} = 63.9% vs. M_{reject} = 64.1%), $t(752) = 0.20$, p = .844).

2.3. Facts Study

We first attempted to extend attribute matching to an objective domain by keeping close to valence and presenting participants a series positive statements (e.g., “More than 2 million dogs were adopted from shelters last year”) and negative statements (e.g., “There were fewer than 100 murders for every 100,000 citizens in Honduras last year”). The statements were a mix of “more than” and “less than” statements. We manipulated the valence of the decision frame by asking some participants if the statements were true (positive frame) and some if the statements were false (negative frame).

Method

Participants. We decided in advance to recruit 400 participants from Amazon Mechanical Turk, and received 404 (M_{age} = 31.9, 51.0% female).

Materials, procedure, design. Participants viewed five positive statements and five negative statements, in a randomized order. We pretested similar statements with 53 Amazon Mechanical Turk participants to be perceived as either positive or negative. Half of participants were asked if the statement was true (yes/no); half were asked if the statement was false. (We instructed participants not to cheat and look up the answers, and designed the statements to be difficult to verify online quickly.) After each trial, as in previous studies, participants reported their confidence and consensus estimates. At the conclusion of the study, participants completed a brief attention check and provided their age and gender. (Pre-registration and data: https://osf.io/jrfbq/?view_only=1e66673c42cf4f419de1400448e8e9ba.)

Results

Fifteen participants (3.7%) failed the attention check and were excluded from analyses.

Confidence. With the remaining participants, we first fit a hierarchical linear model for confidence estimates using clustered standard errors. We defined two Level-1 variables—*frame*, for the valence of decision frame (-0.5 = true, 0.5 = false), and *option*, for the valence of the decision's options (0.5 = less unpleasant, -0.5 = more unpleasant)—and included their interaction. Neither main effect nor the interaction was significant: for the positive statements, participants who were asked if the statements were true were equally confident in their answer (M = 5.53) as the participants asked if the statements were false (M = 5.63), $t(388) = 0.78$, $p = .433$. There was no difference for negative statements either (M_{choose} = 7.22 vs. M_{reject} = 7.48), $t(388) = 0.31$, $p = .759$).

Consensus. We then fit the same model for consensus estimates. Again, neither main effect nor their interaction was significant: for the positive statements, participants who were asked if the statements were true estimated equal consensus in their answer (M = 61.9%) as the participants asked if the statements were false (M = 61.8%), $t(388) = 0.15$, $p = .885$ There was no difference for negative statements either (M_{choose} = 62.9% vs. M_{reject} = 61.9%), $t(388) = 0.92$, $p = .357$).

2.4. Animal Sizes

An early attempt at attribute matching in objective domains focused on animal size. In this study, we showed participants pairs of animals that were either big (e.g., elephant) or small (e.g., termite). We manipulated this attribute of size in the decision frame by asking some participants which animal in the pair was bigger (big frame) and asking others which animal was smaller (small frame). (Pre-registration and data: https://osf.io/6b8tv/?view_only=337e7a4ecc094b179fe06554bab95580.)

Method

Participants. We decided in advance to recruit 400 participants from Amazon Mechanical Turk, and received 405 (M_{age} = 30.8, 53.6% female).

Materials, procedure, design. Participants viewed five pairs of big animal names and five pairs of small animal names, in a randomized order. Half of participants were asked which animal was bigger; half were asked which animal was smaller. After each pair, participants reported their confidence and consensus estimates as in previous studies. At the conclusion of the study, participants completed a brief attention check, provided their age and gender, and gave the names of any animals we included that they did not recognize.

Results

Twenty-seven participants (6.7%) failed the attention check and were excluded from analyses.

Confidence. With the remaining participants, we first fit a hierarchical linear model for confidence estimates using clustered standard errors. We defined two Level-1 variables—*frame*, for the size attribute of the decision frame (-0.5 = which is smaller, 0.5 = which is bigger), and *option*, for the size attribute of the decision's options (0.5 = big, -0.5 = small)—and included their interaction. We found significant effects of both *frame* (b = 0.12, SE = 0.11, p = .015) and *option* (b = 0.39, SE = 0.06, p < .001). The interaction was significant as well (b = -0.55, SE = 0.12, p < .001), however it was not the crossover interaction we had predicted: for the big animals, there was no difference in reported confidence between the two frames ($M_{bigger?}$ = 6.22, $M_{smaller?}$ = 6.24), $t(376)$ = 0.11, p = .913. For the small animals, participants reported greater confidence when asked which is *bigger* (M = 6.11) than which is smaller (M = 5.57) $t(376)$ = 4.24, p < .001, contrary to our predictions. This in itself was unexpected, but, looking closer at the trial-level effects, the results belied greater inconsistencies: for the big animal pairs, three of the five trended in the predicted direction while the remaining two showed significant effects in the opposite direction; the small animal pairs showed a similar conflicting pattern.

Consensus. We then fit the same model for consensus estimates, and found similar results: both main effects were significant (*frame*: b = 2.73, SE = 0.90, p = .003; *option*: b = 4.18, SE = 0.55, p < .001), but the interaction was, again, not the crossover we predicted (b = -3.20, SE = 1.10, p = .004). There was no difference in consensus estimates for the big animal pairs ($M_{bigger?}$ = 66.5%, $M_{smaller?}$ = 65.4%; $t(376)$ = 1.06, p = .289) and the small animal pairs showed the opposite pattern as we predicted ($M_{bigger?}$ = 63.9%, $M_{smaller}$ = 59.6%; $t(376)$ = 4.11, p < .001).

Excluding participants that reported not knowing all the animals included did not meaningfully change the results.

Discussion

The null results from the above two studies initially surprised us. However, we had not yet run Studies 5a and 5b, which suggest that the attribute being matched needs to be naturally, inherently salient for the attribute matching effect to occur. What we believe happened in these two null studies is that the attributes were not nearly salient enough, hence no matching could occur. Additionally, a separate issue may have been that these domains (numeric facts and animal sizes) may have been too obscure for participants to feel comfortable in (note the much lower confidence ratings compared to the first two studies or those in the main text, for example). This concern was one reason why we moved on to (and succeeded with) calorie judgments, a more familiar and accessible domain to most participants.