Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-024-02565-6

BRIEF REPORT

Attribute commensurability and context effects in preferential choice

William M. Hayes[1] · **William R. Holmes**[2] · **Jennifer S. Trueblood**[3]

Accepted: 11 July 2024

Abstract
Context effects in multi-alternative, multi-attribute choice are widely documented, but often elusive. We show that this elusiveness can arise in part from the way that choices are presented. To illustrate this, we use a modeling framework to predict how changes to the format of attribute values, specifically the commensurability of attribute values, influences attention allocation and consequently context effects. Guided by this framework, we show in two online choice experiments (total N = 954 adults) that manipulating the commensurability of attributes leads to different patterns of context effects. Robust attraction and compromise effects are found when attributes are incommensurable (e.g., CPU speed in GHz and RAM memory in GB, or quality ratings on different scales), and mostly null effects occur when attributes are commensurable (e.g., quality ratings on the same scale). Our findings show how the format of choice information can substantially alter the integration of that information and resulting choice patterns.

Keywords Multi-alternative multi-attribute choice · Attention · Evidence accumulation models · Presentation format

Context impacts how people evaluate alternatives and make decisions in both laboratory and real-world settings (Huber et al., 1982; Otto et al., 2022; Simonson, 1989; Tversky, 1972; Wu & Cosguner, 2020). How context affects decisions, and under what circumstances, is less understood. This has led to conflicts in the literature, with seemingly similar studies leading to opposing results (Spektor et al., 2021). We hypothesize that much of the conflict may be a consequence of a failure to account for the role of attention during decision-making. Here, we use a theoretical framework for explicitly and jointly modeling attention and decision processes to identify a key factor, namely attribute commensurability, which can impact attention allocation during deliberation and consequently the appearance of context effects.

✉ William M. Hayes
whayes2@binghamton.edu; wmhayes.psyc@gmail.com

✉ Jennifer S. Trueblood
jstruebl@iu.edu

[1] Department of Psychology, Binghamton University, State University of New York, PO Box 6000, Binghamton, NY 13902, USA

[2] Cognitive Science Program and Department of Mathematics, Indiana University, Bloomington, IN 47405, USA

[3] Department of Psychological and Brain Sciences and Cognitive Science Program, Indiana University, 1101 E 10th St, Bloomington, IN 47405, USA

Early work on contextual sensitivity in multi-alternative, multi-attribute choice focused on how adding an option to a choice set impacted choices for existing options (e.g., the attraction effect; Huber et al., 1982). More recent work has demonstrated that the format in which choice information is presented can influence contextual sensitivity as well. For example, numeric versus pictorial or qualitative attribute formats can lead to different patterns of context effects (Brendl et al., 2023; Frederick et al., 2014; Yang & Lynn, 2014), as can different presentation layouts (Cataldo & Cohen, 2018, 2019). The presence of these types of factors makes it difficult to integrate results of different studies to build a larger understanding of contextual sensitivity in decision-making.

Here we use a theoretical modeling framework to predict how changes to the format of options can induce biases in attention allocation and thus impact context effects. Guided by this framework, we experimentally illustrate how manipulation of a seemingly simple factor, attribute commensurability, can either promote or impede the emergence of context effects in choice.

Attribute commensurability

Consider the purchasing decision between the laptops shown in the top of Fig. 1A. You might compare the CPU speeds of options X and Y, followed by the RAM sizes of options Y

Published online: 16 September 2024

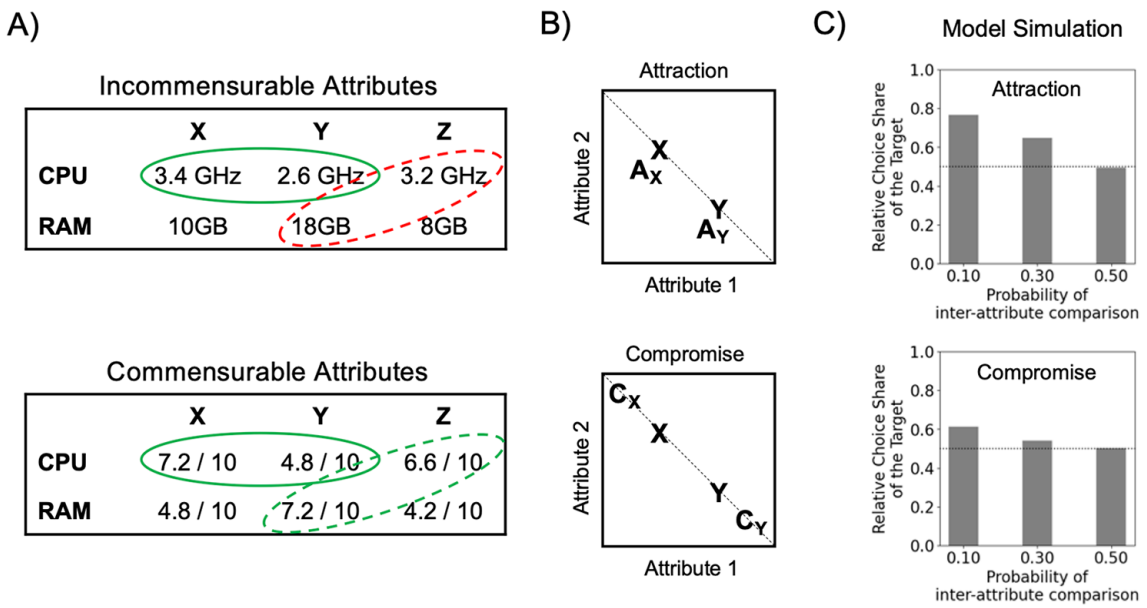

Fig. 1 Attribute commensurability and model predictions. **A** Example of choice scenarios with incommensurable (*top*) and commensurable attributes (*bottom*). The *solid ovals* depict intra-attribute comparisons, and the *dashed ovals* depict inter-attribute comparisons. **B** Choice contexts used in model simulations for producing the attraction and compromise effects. The *dotted line* represents the line of indifference, assuming equal weighting of the attributes. The two core options $X = (0.3, 0.7)$ and $Y = (0.7, 0.3)$ are paired with one of two decoys designed to favor either X or Y. The attraction decoys $A_X = (0.2, 0.6)$ and $A_Y = (0.6, 0.2)$ are dominated by the target option (denoted with subscripts), but not by the competitor. The compromise decoys $C_X = (0.1, 0.9)$ and $C_Y = (0.9, 0.1)$ are placed so that the target will have intermediate values in the choice set. Both decoys are expected to increase the relative preference for the target over the competitor. **C** Proportion of times the model chose the target over the competitor as a function of the probability of inter-attribute comparisons. Results are based on 1000 simulations. The parameter values used for the simulations were $\alpha = 1$, $\lambda = 5$, $w = 0.5$, $b = 0.2$

and *Z*. However, GB and GHz are incommensurate measures, and therefore the CPU speed of one option is unlikely to be compared with the RAM size of another option.

Now consider the scenario shown in the bottom of Fig. 1A where the same laptops are presented with quality ratings for CPU and RAM. The common scale makes it easier to compare across attributes, which should in turn increase the probability of making inter-attribute comparisons (Evangelidis & van Osselaer, 2019; Simonson et al., 2013). This small change in presentation format can thus change the space of potential evaluations that are performed over the course of a decision. How might this affect the resulting choices?

The distinction between commensurable and incommensurable attributes is important to many real-world decisions. When shopping online, for example, people often view product specifications that are on different scales (e.g., CPU speed, memory size, screen size, weight) and product evaluations that are on the same scale (e.g., reviewer ratings). A recent study found that context effects were diminished when attributes were expressed as ratings on a common scale (Banerjee et al., 2024). However, a computational process-based account of this effect is currently lacking. Here, we offer a model-based explanation of the moderating effect of attribute commensurability that relies on the allocation of attention to different types of comparisons during deliberation. In contrast to existing process models, which assume that decision-makers only attend to intra-attribute comparisons (Roe et al., 2001; Trueblood et al., 2014; Usher and McClelland, 2004; for a review, see Trueblood, 2022), our model allows for both intra- and inter-attribute comparisons. Importantly, the model predicts that attending to different types of comparisons results in different patterns of context effects. Below we describe the theoretical framework and its predictions.

Theoretical approach: Comparison sampling model

In choice sets with many options containing several attributes, attention determines which attributes are attended to and how they are evaluated. In the present work, we use a theoretical framework (first presented in Trueblood et al. 2022) that explicitly incorporates attention effects into a sequential sampling model of multi-alternative, multi-attribute choice. This framework builds on ideas from previous models, particularly the idea that attention switches stochastically between different comparisons over time ((Noguchi & Stewart, 2018; Roe et al., 2001; for a review, see Trueblood, (2022)). However, it differs in that it explicitly models the probabilities of switching to different comparisons, allowing it to account for a range of different factors that might bias attention,

such as presentation format. The framework is premised on three basic assumptions: (1) Information from the choice set is sequentially sampled over time until sufficient support for one alternative triggers a decision, (2) evidence (or preference) for an alternative is sampled by comparing it to some referent (e.g., another alternative), (3) attention determines the sequence of comparisons that occur over time.

In this framework, a decision is composed of a discrete and finite sequence of comparisons over time, each determining what information is being attended to and accumulated. Attention is the process that determines which comparisons are made. This is a flexible, general framework that allows the researcher to incorporate and compare different assumptions about how attention might modulate comparisons (see Trueblood et al. 2022). For brevity, we will assume that people perform pairwise comparisons (i.e., comparing one alternative to another; Noguchi and Stewart, 2014; Tversky and Simonson, 1993) and that more time is spent on comparisons where alternatives are more similar (Trueblood et al. 2014). See Supplemental Note 1 for further details.

Model predictions

We hypothesize that attribute commensurability should influence the probability of attending to inter-attribute comparisons (Evangelidis & van Osselaer, 2019). Most theoretical accounts of multi-attribute decision-making assume that attributes are evaluated independently. Our theoretical framework relaxes this assumption, modeling the probability of attending to inter-attribute comparisons as a free parameter, p_{inter}. To generate predictions for the effects of inter-attribute comparison on choice, we simulated the model in two artificial choice contexts, each composed of three options. One context was designed to produce the attraction effect (Huber et al., 1982), which occurs when the presence of a "decoy" option that is dominated by one of the other options increases preference for the dominating option (Fig. 1B, top). The other context was designed to produce the compromise effect (Simonson, 1989), which occurs when a decoy option that is extremely good on one attribute and extremely poor on another increases preference for the option with intermediate values on both attributes (Fig. 1B, bottom). The option that the decoy is designed to favor is called the "target," and the other option is called the "competitor."

As shown in Fig. 1C, the model predicts a preference for the target when p_{inter} is low (i.e., standard attraction and compromise effects), which should be the case when attributes are incommensurable. As p_{inter} increases, which might occur with commensurable attributes, the context effects diminish. This happens because the intra-attribute comparisons that favor the target alternative are increasingly cancelled out by inter-attribute comparisons that favor the competitor (see Supplemental Note 1 for a demonstration). When $p_{inter} = 0.50$, the model predicts null effects. Thus, according to our model, the attraction and compromise effects should be restricted to choice environments where the probability of inter-attribute comparisons is low.

The present study was designed to test this prediction by experimentally manipulating attribute commensurability. Participants made repeated ternary choices between hypothetical consumer products, where one of the options on each trial was a decoy designed to favor one of the other two options. We tested both attraction and compromise decoys. In the first experiment, participants were randomly assigned to one of three conditions. In one condition, the attributes were separate product features (e.g., CPU speed and RAM size) with incommensurable values (e.g., GHz and GB). In the other two conditions, the attribute values were expressed on a common rating scale but the attributes themselves were either separate features or overall quality ratings. The second experiment was designed to disentangle the effects of attribute type and attribute commensurability using a 2x2 factorial design. Based on the predictions of our modeling framework, we hypothesized that the attraction and compromise effects would be larger in the conditions with incommensurable attribute values.

Experiment 1

Experiment 1 tested whether attribute presentation format moderates the attraction and compromise effects. Participants in the Incommensurable / Separate Features condition made choices between laptops described by CPU speed (GHz) and RAM size (GB), or washing machines described by price ($) and capacity (cubic ft.). The Commensurable / Separate Features condition used the same attributes, but the values were presented as ratings from "expert reviewers" on a common scale. To account for the possibility that unequal attribute weighting could make it more difficult to compare across attributes (e.g., if someone weights CPU speed much more than RAM), the Commensurable / Overall Evaluations condition used the same rating values but with the attribute labels "Reviewer 1" and "Reviewer 2." The ratings were meant to represent the overall evaluations of two expert reviewers. Because there is no a priori reason to weight Reviewer 1's rating more than Reviewer 2's (or vice versa), we reasoned that these attributes may be even easier to compare. On the other hand, if the commensurability of the attribute values is all that matters, behavior in the two commensurable conditions should be very similar.

Method

Participants

For the attraction effect, 360 participants (women = 223, men = 134, non-binary = 2, unreported = 1; age: M = 43.55, SD = 13.26) were recruited from Amazon Mechanical Turk using CloudResearch and randomly assigned to one of the three conditions (see below). The sample size was chosen to be larger than those in past studies that tested moderators of context effects using a between-subjects design (e.g., Cataldo and Cohen, 2021; Trueblood et al. 2022). Based on our preregistered exclusion criteria (AsPredicted #111511), data for 41 participants were excluded for failing more than 1/3 of the catch trials. The final sample included 102, 104, and 113 participants in the Incommensurable/Separate, Commensurable/Separate, and Commensurable/Overall conditions, respectively.

For the compromise effect, 358 participants were recruited from Amazon Mechanical Turk using CloudResearch (women = 186, men = 171, non-binary = 1; age: M = 42.46, SD = 12.96). Our preregistered sample size was 360, but two participants were lost due to a data recording error. Based on our preregistered exclusion criteria, data for 59 participants were excluded for failing more than 1/3 of the catch trials. The final sample included 97, 92, and 110 participants in the Incommensurable/Separate, Commensurable/Separate, and Commensurable/Overall conditions, respectively.

The compromise effect condition was run concurrently with the attraction effect condition, but participants were prevented from participating in both. Participants were compensated $1.25 for completing the study. The experiment was approved by the Institutional Review Board at Indiana University.

Materials

Participants encountered several choices between laptops and washing machines. Example attraction effect choice sets are shown in Fig. 2.

Each choice consisted of two focal options (X and Y) and a decoy option that favored one of the two focal options. The attribute values varied across trials. In the Incommensurable/Separate condition, on laptop trials, X had a CPU speed ranging between 3.0 GHz and 3.9 GHz, and a RAM size ranging between 6GB and 15GB. Y had a CPU speed ranging between 2.2 GHz and 3.1 GHz, and a RAM size ranging between 14GB and 23 GB. On washing machine trials, X had a price ranging between $500 and $590, and a capacity ranging between 2.5 and 3.4 cubic ft. Y had a price ranging between $660 and $750, and a capacity ranging between 4.1 and 5.0 cubic ft. The attribute values for X and Y were selected so that X was always 0.8 GHz faster (laptops) or $160 cheaper (washing machines) than Y, but with 8GB less RAM or 1.6 cubic ft. less capacity. In the Commensurable/Separate condition, the rating for X on CPU (laptops) or price (washing machines) ranged between 5.9 and 8.7 points, while the rating for Y ranged between 3.5 and 6.3 points. The ratings for RAM (laptops) and capacity (washing machines) were flipped such that X was always rated 2.4 points higher

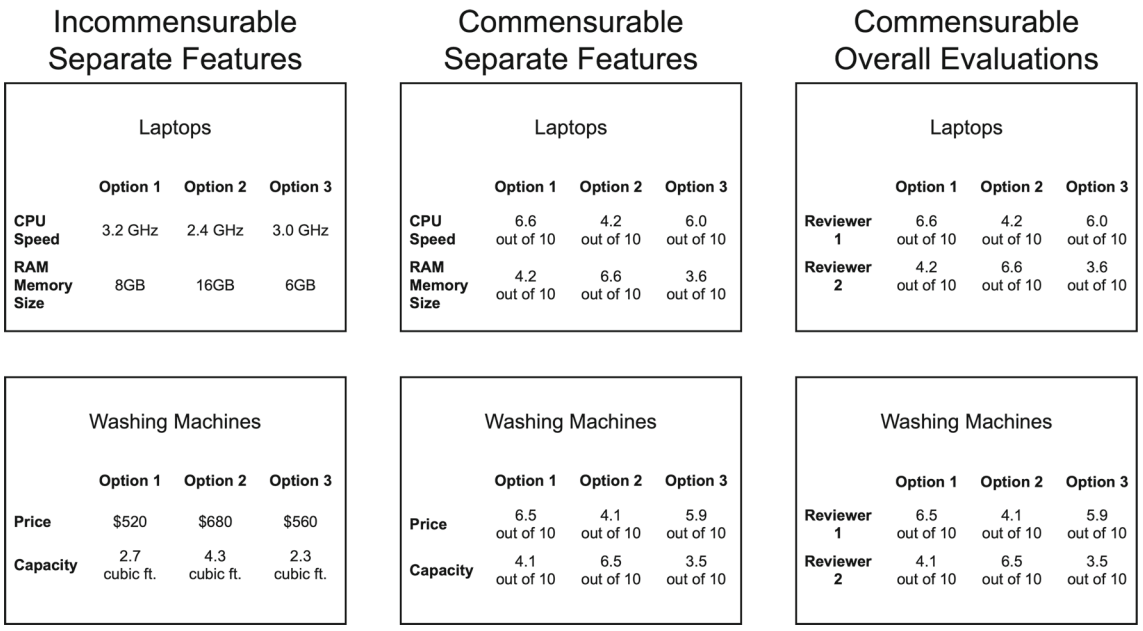

Fig. 2 Experiment 1 conditions. Example choice scenarios in the three conditions. Each scenario depicts an attraction effect trial where the order of the alternatives from left to right is X, Y, A_X

on CPU or price, but 2.4 points lower on RAM or capacity compared to *Y*. The same rating values were used in the Commensurable/Overall condition, where *X* was always rated 2.4 points higher by Reviewer 1 than *Y*, but 2.4 points lower by Reviewer 2 for both laptops and washing machines. A full list of the attribute values can be found in Supplemental Tables 6 and 7.

The attraction decoys were designed to be similar, but inferior to their corresponding target option. The attribute values for A_X and A_Y were created by shifting away from the target option's values by 25% of the range between *X* and *Y* (e.g., A_X was always 0.2 GHz or 0.6 rating points lower on CPU and 2 GB or 0.6 rating points lower on RAM than *X*).

The compromise decoys were constructed to be better than the target on its better attribute, but worse than the target on its worse attribute. The attribute values for C_X and C_Y were created by shifting away from the target option's values by 50% of the range between *X* and *Y* (e.g., C_X was always 0.4 GHz or 1.2 rating points higher on CPU and 4 GB or 1.2 rating points lower on RAM than *X*).

In total, there were 40 trials with *X* as the target option and 40 trials with *Y* as the target option, half with laptops and half with washing machines. In addition to these test trials, there were 40 catch trials with a single dominating option. The catch trials included choices between laptops and washing machines as well as airline tickets (price / flight duration), auto loans (monthly payment / loan term), and cars (fuel economy / ride quality), in order to vary the stimuli. Thus, there were 120 trials in total (80 test and 40 catch trials).

Procedure

The instructions at the beginning of the experiment stated that participants would encounter several choice scenarios, each with three options, and that their task was to choose the option that they prefer. Following this, participants read a description of the choice categories and attributes. Participants in the Commensurable/Separate condition were told that the options were rated on two different attributes by expert reviewers. Participants in the Commensurable/Overall condition were told that each option was rated by two expert reviewers. In both cases, the instructions stated that the ratings range from 0 to 10, where 10 is the best possible value. Participants completed three practice trials prior to starting the actual experiment.

At the beginning of each trial, participants were presented with three options along with their attributes. The options and their attributes were displayed in a table with the two attributes in different rows and the three options in different columns. The column labels were always "Option 1," "Option 2," and "Option 3," in that order. However, the order of the underlying options (i.e., *X*, *Y*, and the decoy) was randomized on each trial. The order of the attributes was also randomized on each trial except for the Commensurable/Overall condition, where the ratings from "Reviewer 1" always appeared in the first row and the ratings from "Reviewer 2" always appeared in the second row. Participants used the 1, 2, and 3 keys to select an option. There was no feedback provided after a choice. The order of the trials was randomized with the constraint that participants never saw two test trials with laptops, or two test trials with washing machines, on consecutive trials.

Data analysis

To quantify context effects, we calculated the *relative choice share for the target* (RST) for each participant (equal-weights version; see Katsimpokis et al. 2022). The RST gives the proportion of times that the target, or the option favored by the decoy, was selected over the competitor. For example, if a person almost always chooses *X* when the decoy targets *X* and *Y* when the decoy targets *Y*, their RST value should be close to 1.0. On the other hand, a person who strongly prefers either *X* or *Y* regardless of the decoy, or a person who is indifferent between them, should have an RST value close to 0.5.

The equal-weights version of RST is meant to protect against biased inferences that can result when the total number of target and competitor selections differs across contexts where the decoy targets *X* versus *Y* (Katsimpokis et al., 2022). This can happen when the decoy is strongly preferred in one of the two contexts, e.g., in compromise effect choice sets where the decoy may be an attractive option. Let T_X and C_X denote the total number of target and competitor selections when the decoy favors *X*, and let T_Y and C_Y denote the total number of target and competitor selections when the decoy favors *Y*. Then the equal-weights RST is defined as follows:

$$RST_{EW} = 0.5 \cdot \left(\frac{T_X}{T_X + C_X} + \frac{T_Y}{T_Y + C_Y} \right). \quad (1)$$

RST > 0.5 indicates the presence of a context effect, RST $= 0.5$ a null effect, and RST < 0.5 a reversed context effect.

We used a hierarchical Bayesian approach to model the RST values in each condition. Following previous studies (Trueblood, 2015; Trueblood et al., 2015), the model assumes that the number of target selections in a particular context follows a binomial distribution:
$T \sim Binomial(\theta, T + C)$, where θ is the probability of selecting the target over the competitor. To accommodate the equal-weights version of RST, separate θ parameters are estimated for contexts where the decoy favors *X* and contexts where the decoy favors *Y*: $T_X \sim Binomial(\theta_X, T_X + C_X)$ and $T_Y \sim Binomial(\theta_Y, T_Y + C_Y)$ (Katsimpokis et al., 2022). In total, four θ parameters were estimated for each

person: one for laptops when the decoy favors *X*, one for laptops when the decoy favors *Y*, one for washing machines when the decoy favors *X*, and one for washing machines when the decoy favors *Y*. The person-specific θ parameters were assumed to be drawn from group-level beta distributions with mean μ and concentration κ. The priors for the hyper-parameters were chosen to be relatively vague: $\mu \sim Beta(1, 1)$ and $\kappa \sim |N(0, 10)|$. We modeled the three attribute conditions separately and obtained a group-level estimate of the equal-weights RST in each condition by averaging the posterior distributions for μ_X, the mean probability of selecting the target when the decoy favors *X*, and μ_Y, the mean probability of selecting the target when the decoy favors *Y*: $\mu_{RST} = 0.5 \cdot (\mu_X + \mu_Y)$. A graphical diagram of the model can be found in Supplemental Note 2.

In accordance with our preregistration, we also tested a hierarchical Bayesian model for the traditional version of RST (Trueblood, 2015; Trueblood et al., 2015), and a multinomial logit model that measures overall contextual sensitivity. The results from these analyses were similar to the results using the equal-weights RST and are available in Supplemental Note 2.

Results and discussion

The results are summarized in Fig. 3. For both types of decoys, the mean RST values for laptops and washing machines were well above 0.50 in the Incommensurable/Separate condition (context effect), but very close to 0.50 in both Commensurable conditions (null effect). For the attraction decoy, there were a small number of RST values close to zero in the Commensurable/Overall condition, consistent with a reverse attraction effect. There was substantial individual variability in response to the compromise decoy, with some participants exhibiting strong compromise effects, some showing null effects, and some showing strong reverse effects. However, the majority of individual RST values in the Incommensurable condition were above 0.50.

Table 1 shows the estimated posterior mean and the 95% highest posterior density (HPD) interval for the mean RST (μ_{RST}) in each experimental condition, derived from the hierarchical Bayesian model. Separate estimates for laptops and washing machines are shown, along with overall estimates. The posterior distribution for the overall estimate was computed by averaging the posterior distributions of the category-specific estimates: $\mu_{RST,overall} = 0.5 \cdot (\mu_{RST,laptops} + \mu_{RST,washing})$. HPD intervals entirely above 0.50 indicate a positive context effect, while HPD intervals containing 0.50 indicate a null effect.

A robust attraction effect was observed in the Incommensurable condition for both laptops and washing machines, as well as averaging across product categories. There was no evidence for the attraction effect in the Commensurable conditions, as all of the HPD intervals included 0.50. Similarly, there was an overall compromise effect in the Incommensurable condition and an overall null effect in the Commensurable conditions. Results were similar using frequentist statistics (Supplemental Tables 8 and 9).

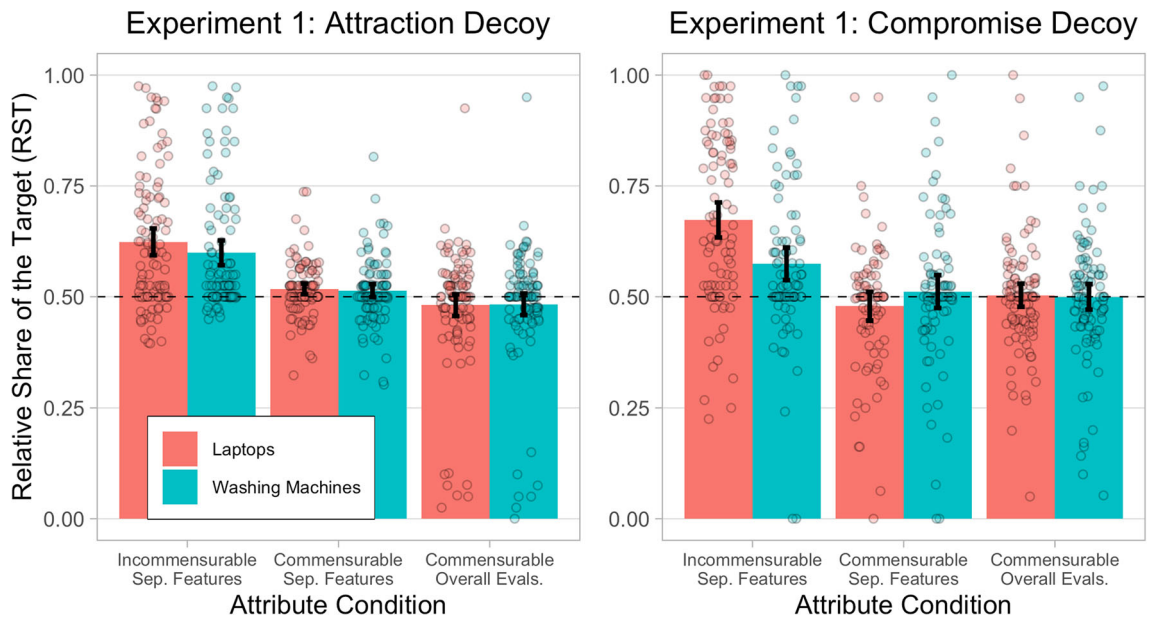

Fig. 3 Experiment 1 results. Empirical RST values by choice category and attribute condition (equal-weights version; Eq. 1). Individual estimates are shown as *points*. For the attraction (compromise) decoy, three points (35 points) were excluded because the RST was undefined. *Error bars* represent 95% confidence intervals

Table 1 Group-level equal-weights RST estimates from the hierarchical Bayesian model (μ_{RST})

Condition	Laptops		Washing Machines		Overall	
	Mean	95% HPD	Mean	95% HPD	Mean	95% HPD
Experiment 1: Attraction Decoy						
Incommensurable/Separate	0.623	[0.57, 0.68]	0.742	[0.58, 0.92]	0.682	[0.60, 0.77]
Commensurable/Separate	0.531	[0.47, 0.60]	0.523	[0.44, 0.60]	0.527	[0.48, 0.58]
Commensurable/Overall	0.485	[0.45, 0.53]	0.482	[0.44, 0.52]	0.483	[0.46, 0.51]
Experiment 1: Compromise Decoy						
Incommensurable/Separate	0.679	[0.62, 0.74]	0.628	[0.43, 0.83]	0.654	[0.54, 0.76]
Commensurable/Separate	0.459	[0.40, 0.52]	0.494	[0.44, 0.55]	0.477	[0.44, 0.52]
Commensurable/Overall	0.498	[0.46, 0.54]	0.490	[0.45, 0.53]	0.494	[0.47, 0.52]

We fit the comparison sampling model to the individual choice-RT data using Bayesian methods (test trials only; see Supplemental Note 1 for details). In the attraction (compromise) condition, the model explained $> 95\%$ ($> 81\%$) of the variance in choice proportions and $> 94\%$ ($> 94\%$) of the variance in mean RTs across participants (Supplemental Note 4). The model also captured relationships between preference and deliberation time, including the increasing target preference and decreasing competitor preference with longer deliberation in the Incommensurable/Separate condition, and the absence of these trends in the Commensurable/Overall condition (Supplemental Figures 7 and 8). For both decoys, the median estimate of the probability of attending to inter-attribute comparisons (p_{inter}) was lowest in the Incommensurable/Separate condition and highest in the Commensurable/Overall condition (Fig. 4).

In summary, the results support the predictions of the comparison sampling model. Aggregate attraction and compromise effects were observed in the condition with incommensurable attributes. In the two conditions with commensurable attributes, the effects disappeared. These results may be due to a greater probability of attending to inter-attribute comparisons when the attributes are commensurable.

Experiment 2

In the previous experiment, aggregate context effects were observed when the attributes were on different scales (e.g., CPU speed in GHz and RAM size in GB), but not when the attributes were presented on a common rating scale. Our results are consistent with a recent study (Banerjee et al., 2024), and suggest that the commensurability of the attribute values was the primary driver. Another possibility is that context effects diminish whenever the attributes are presented in a ratings format, regardless of whether the rating scales are commensurable or not. Because the previous experiments and the Banerjee et al. (2024) study lack a condition with incommensurable rating scales, however, they cannot be used to answer this question.

Further, different *types* of attributes were presented in the three conditions: either separate features (e.g., CPU and RAM) or overall evaluations (ratings from separate reviewers). As mentioned previously, the purpose of using overall evaluations from "Reviewer 1" and "Reviewer 2" was to minimize the impact of unequal attribute weighting, which could make it difficult to compare across attribute dimensions even if the values are commensurable. However, because attribute type and value commensurability were manipulated using a non-factorial design, the previous experiment does not cleanly disentangle the effects of these two factors.

Experiment 2 was designed to address these limitations. Attribute type and commensurability were manipulated in a 2x2 between-subjects design. Participants made hypothetical choices between laptops where the attributes were either separate product features (CPU speed and RAM size) or overall evaluations (ratings from "Reviewer 1" and "Reviewer 2"). The attribute values were either incommensurable or commensurable (percentages). Further, the attributes in all four conditions were bounded with explicit ranges.

We predicted that the attraction and compromise effects (manipulated within-subjects) would be observed in the conditions with incommensurable attribute values, but not in the conditions with commensurable values (AsPredicted #154254). We did not predict an effect of attribute type or an interaction.

Method

Participants

A total of 346 participants were recruited from Prolific (women = 170, men = 169, trans women = 1, trans men = 3, other = 2, prefer not to say = 1; age: M = 40.53, SD = 13.08). Our preregistered sample size was 400, but 54 partic-

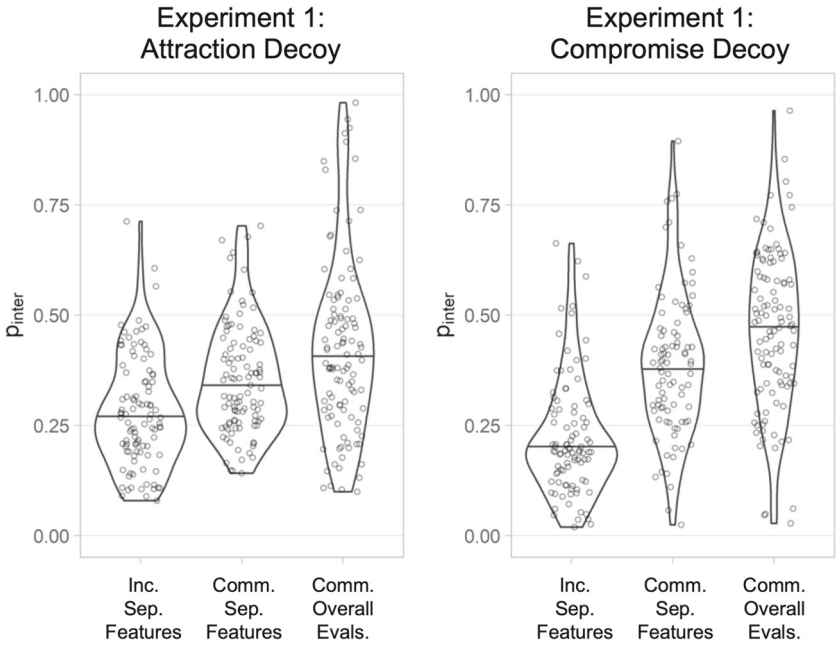

Fig. 4 Model-estimated probability of attending to inter-attribute comparisons in Experiment 1. *Points* show the posterior means of the individual-level parameters. *Horizontal lines* show the median estimates for each condition

ipants were lost due to a data recording error. Following our preregistered exclusion criteria, data for ten participants were excluded for failing more than 1/3 of the catch trials. The final sample included $n = 94$ in the Incommensurable/Separate condition, $n = 80$ in the Commensurable/Separate condition, $n = 76$ in the Incommensurable/Overall condition, and $n = 86$ in the Commensurable/Overall condition. Participants were paid $4.00 for completing the study. The experiment was approved by the Institutional Review Board at Binghamton University.

Materials

Participants made repeated choices between laptops. Each choice set consisted of two focal options (X and Y) and a decoy option that targeted one of the two focal options. In the separate features conditions, the attributes were CPU speed and RAM memory size. In the overall evaluation conditions, the attributes were ratings from Reviewer 1 and Reviewer 2. The attribute values were either incommensurable or commensurable, and they varied across trials. Example choices in each condition are shown in Fig. 5.

In the incommensurable conditions, the value of X on the first attribute (CPU speed or Reviewer 1) ranged between 3.0 and 4.8 in increments of 0.2 (ten levels), while the corresponding value on the second attribute (RAM size or Reviewer 2) ranged between 6 and 24 in increments of 2 (ten levels). For each option X_i (i = 1,...,10), there was a corresponding option Y_i constructed by subtracting 0.8 units from the first attribute value and adding 8 units to the second attribute value. Thus, for example, when X had attribute values (3.0, 6), Y had attribute values (2.2, 14), and when X = (4.8, 24), Y = (4.0, 32). The attribute values for the attraction decoys A_X and A_Y were lower than the target option's values by 25% of the difference between X and Y on the corresponding attribute. The compromise decoys C_X and C_Y were higher than the target on its better attribute and lower than the target on its worse attribute by an amount equal to 50% of the difference between X and Y. For example, the choice set with focal options X = (4.0, 16) and Y = (3.2, 24) would have corresponding attraction decoys A_X = (3.8, 14) and A_Y = (3.0, 22), and compromise decoys C_X = (4.4, 12) and C_Y = (2.8, 28). Note that across all choice sets, the values on the first attribute (CPU speed or Reviewer 1) ranged between 1.8 and 5.2, while the values on the second attribute (RAM size or Reviewer 2) ranged between 2 and 36.

The attribute values in the commensurable conditions were created by converting the incommensurable values to percentages using the formula $\% = [(x - \min)/(\max - \min)] * 100$, and rounding to the nearest integer. For example, the choice set with focal options X = (4.0, 16) and Y = (3.2, 24) would become X = (65%, 41%) and Y = (41%, 65%). The decoys in this example would become A_X = (59%, 35%), A_Y = (35%, 59%), C_X = (76%, 29%), and C_Y = (29%, 76%).

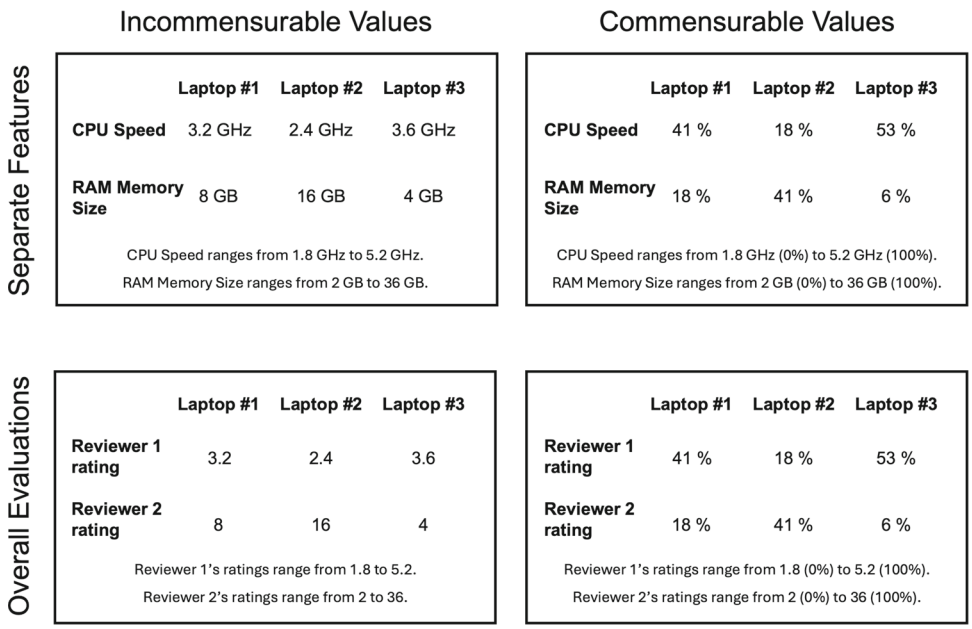

Fig. 5 Experiment 2 conditions. Example choice scenarios in the four conditions. Each scenario depicts a compromise effect trial where the order of the alternatives from left to right is X, Y, C_X. The order of the options varied randomly across trials, but the order of the attributes remained fixed. Attribute values also varied across trials. Ranges for the attributes were explicitly presented at the bottom of the screen in all four conditions

Procedure

In the separate features conditions, the instructions defined CPU speed and RAM memory size and stated that a higher CPU speed / RAM size is always better. Additionally, participants were informed that CPU speed (RAM size) is measured in gigahertz (gigabytes) and that the laptops they would be seeing have CPU speeds (RAM sizes) ranging from 1.8 to 5.2 GHz (2–36 GB). For the commensurable condition, we added the lines: "For convenience, the values have been converted to percentages: 0% = worst possible CPU speed (RAM size), 100% = best possible CPU speed (RAM size)."

The instructions in the overall evaluation conditions were similar. Participants were told that the laptops had been given overall quality ratings by two expert reviewers. They were told that Reviewer 1 gave each laptop a score ranging from 1.8 to 5.2, and that Reviewer 2 gave each laptop a score ranging from 2 to 36 (higher scores being better). For the commensurable condition, we added the line: "For convenience, both reviewers' ratings have been converted to percentages: 0% = worst possible score, 100% = best possible score."

Each trial included three alternatives and their attributes arranged in a 2 (attributes) x 3 (alternatives) table (Fig. 5). The ordering of the underlying options (i.e., X, Y, and the decoy) was randomized on each trial. In the separate features conditions, CPU speed appeared in the first row and RAM memory size appeared in the second row. In the overall evaluation conditions, Reviewer 1's rating appeared in the first row and Reviewer 2's rating appeared in the second row. In the Separate-Incommensurable condition, the units (GHz and GB) appeared along with the values in the table. At the bottom of the screen was a note that contained the ranges for both attributes. In the commensurable values conditions, the ranges were stated in the original units with the percentages in parentheses. The note also contained the statements,"A higher [CPU speed / RAM size / rating] is always better." Participants used the 1, 2, and 3 keys to make a choice (self-paced).

There were 40 attraction and 40 compromise trials, half of each with X as the target option and the other half with Y as the target. Because there were ten unique choice sets for every combination of decoy type and target, each unique choice set appeared twice. In addition to the 80 test trials, there were 40 catch trials with a single dominating option (120 trials in total). The order of trials was randomized for each participant.

Data analysis

We used the same hierarchical Bayesian beta-binomial model for the equal-weights RST that was used in the previous experiment. The four experimental conditions were modeled separately. In total, four θ parameters were estimated for each person: one for attraction trials with A_X, one for attraction trials with A_Y, one for compromise trials with C_X, and one for compromise trials with C_Y. Group-level estimates of the equal-weights RST for each decoy type were used to quantify aggregate attraction and compromise effects.

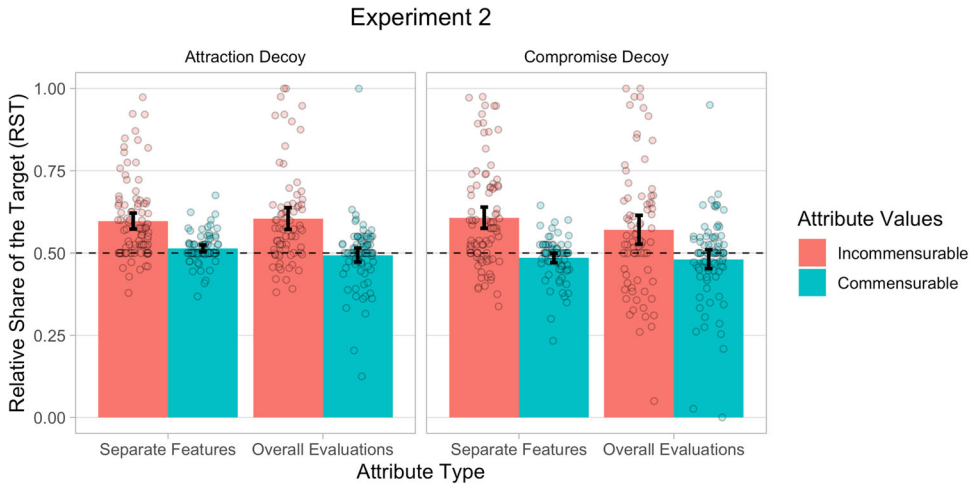

Fig. 6 Experiment 2 results. Empirical RST values by decoy, attribute type, and attribute value conditions (equal-weights RST). Individual estimates are shown as points (six points were excluded because the RST was undefined). *Error bars* represent 95% confidence intervals

In accordance with our preregistration, we also tested a hierarchical Bayesian model for the traditional version of RST (Supplemental Note 2). Although the results for the compromise effect differed somewhat from the equal-weights RST, they should be interpreted with caution due to the traditional RST's susceptibility to bias (Katsimpokis et al. 2022). In the following, we restrict focus to the equal-weights RST.

Results and discussion

The results are summarized in Fig. 6. The mean equal-weights RST values were above 0.50 in the incommensurable conditions, but very close to 0.50 in the commensurable conditions. Attribute type (separate features or overall evaluations) appeared to have little to no effect, and there was no interaction. The pattern of results was very similar for the attraction and compromise decoys. At the individual participant level, strong context effects (RST $>$ 0.75) were much more frequent in the incommensurable conditions.

Table 2 shows the posterior means and 95% HPD intervals for the mean RSTs (μ_{RST}) in each experimental condition. The results confirmed the presence of context effects in the incommensurable conditions, and null effects in the conditions with commensurable values. Results were similar using frequentist statistics (Supplemental Table 10).

We fit the comparison sampling model to the individual choice-RT data following the same procedures from the prior experiment (Supplemental Note 1). The model explained between 74% and 84% of the variance in choice proportions and between 88% and 95% of the variance in mean RTs across the four conditions (Supplemental Note 5). The p_{inter} estimates tended to be lower in the conditions with incommensurable attribute values (Fig. 7).

In summary, aggregate attraction and compromise effects occurred when attributes were incommensurable, but not when they were commensurable (i.e., percentages). This was true regardless of whether the attributes represented separate features of the alternatives (CPU speed and RAM size) or evaluations of overall quality (reviewer ratings).

General discussion

The elusiveness of context effects in decision-making has led researchers to question their importance (Frederick et al., 2014; Huber et al., 2014; Trendl et al., 2021; Yang & Lynn, 2014). However, rather than diminishing the importance or usefulness of context effects, we believe these findings high-

Table 2 Group-level equal-weights RST estimates (μ_{RST}) in Experiment 2

Attribute type	Attribute values	Attraction		Compromise	
		Mean	95% HPD	Mean	95% HPD
Separate features	Incommensurable	0.598	[0.54, 0.66]	0.581	[0.53, 0.63]
	Commensurable	0.512	[0.39, 0.64]	0.473	[0.35, 0.61]
Overall evaluations	Incommensurable	0.608	[0.56, 0.66]	0.558	[0.51, 0.61]
	Commensurable	0.505	[0.45, 0.56]	0.466	[0.41, 0.53]

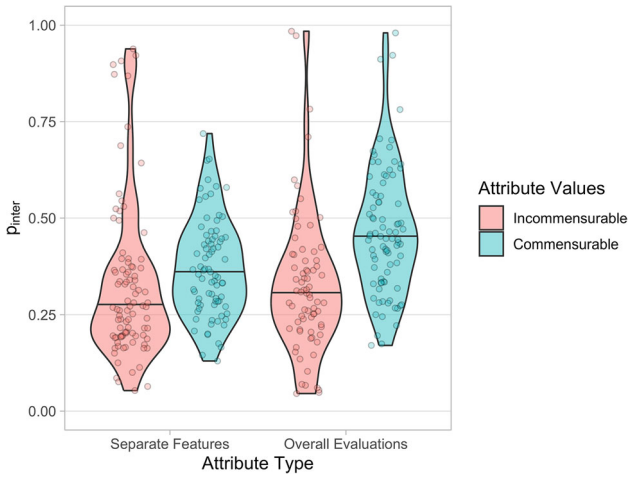

Fig. 7 Model-estimated probability of attending to inter-attribute comparisons in Experiment 2. *Points* show the posterior means of the individual-level parameters. *Horizontal lines* show the median estimates for each condition

light the need for theories that can explain how features of the choice environment give rise to different patterns of behavior. Building on recent work (Cataldo & Cohen, 2021; Trueblood et al., 2022), we propose that presentation format influences the allocation of attention during deliberation, which in turn leads to a range of possible decision outcomes. We focused on a novel aspect of presentation format, *attribute commensurability*, and showed how we can leverage a recently developed modeling framework to understand its impact on the attraction and compromise effects.

Whether the attraction and compromise effects emerge appears to partly depend on attribute commensurability. We observed robust aggregate-level context effects when the attributes were incommensurable (e.g., CPU speed in GHz and RAM size in GB), but null effects when the attributes were presented on a common rating scale and therefore easier to compare. Experiment 2 confirmed that value commensurability is the key moderator: When attributes were presented as ratings but on *different* scales, context effects still emerged.

We used the comparison sampling model, which explicitly accounts for the role of attention in multi-alternative, multi-attribute choice, to understand the influence of attribute commensurability on decision patterns (Trueblood et al., 2022). Fitting the model to the choice-RT data indicated that attribute commensurability modulated the probability of attending to inter-attribute comparisons during the deliberation process. When this probability is low, the model produces attraction and compromise effects, but the effects diminish as greater attention is allocated to inter-attribute comparisons. We hypothesized that inter-attribute comparisons would be less frequent when attributes are incommensurable. In both experiments, the parameter estimates differed between conditions in a manner consistent with our hypothesis.

To establish that the comparison sampling model is the best model for our data, we would need to systematically compare it against other existing models. Here, our goal was simply to present the model as one plausible account for why attribute commensurability moderates context effects and use it to generate testable predictions. Although the data aligned with these predictions, we acknowledge that other cognitive mechanisms could have produced similar results. It is possible that some participants used entirely different strategies when the attributes were commensurable; for example, they may have simply summed or averaged the attribute values to generate a single subjective value for each alternative (i.e., a weighted-additive strategy). Future studies will need to investigate this further. In any case, switching to a weighted-additive strategy would still be expected to lead to a reduction in context effects when attributes are commensurable.

Noguchi and Stewart (2014) analyzed eye movement data in a choice task with incommensurable attributes and found that between-alternative transitions were more frequent than within-alternative transitions (see also Cataldo and Cohen, 2019). While this is consistent with the comparison sampling model, the authors noted that between-alternative, between-attribute transitions, or what we would call inter-attribute transitions, were excluded from the analysis. Incorporating process tracing measures will provide a stronger test of the model and the proposed attention mechanisms for explaining the moderating effects of presentation format.

In summary, our results have important implications for the ongoing debate about the elusiveness of context effects in decision-making. We demonstrate that the elusive nature of context effects may be partially explained by the dynamics of attention during deliberation. External features of the presentation format can influence the allocation of attention to different types of attribute comparisons, resulting in different patterns of choice behavior. By deepening our understanding of the cognitive processes underlying multi-alternative, multi-attribute choice, we can form more accurate predictions for when context effects should occur and when they should not.

Open practices statement

Both experiments were preregistered on AsPredicted.org (Experiment 1: #111511, Experiment 2: #154254). Data and code are available on the Open Science Framework (https://osf.io/htm7b/).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13423-024-02565-6.

Funding This research was supported by NSF grants 1846764/2305559 and 2242962, and by Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute.

Availability of Data and Materials Data and code are available on the Open Science Framework (https://osf.io/htm7b/).

Declarations

Conflicts of Interests/Competing Interests The authors have no conflicts of interest or competing interests to disclose.

Ethics Approval Experiment 1 and Experiment 2 were approved by the Institutional Review Boards at Indiana University and Binghamton University, respectively.

Consent to Participate Informed consent was obtained from all individual participants included in the study.

Consent for Publication All participants consented to the publication of their de-identified data.

References

Banerjee, P., Rakshit, K., Mishra, S., & Masters, T. (2024). Attribute ratings and their impact on attraction and compromise effects. *Marketing Letters,* 1–12.

Brendl, C. M., Atasoy, Ö., & Samson, C. (2023). Preferential attraction effects with visual stimuli: The role of quantitative versus qualitative visual attributes. *Psychological Science, 34*(2), 265–278.

Cataldo, A. M., & Cohen, A. L. (2018). Reversing the similarity effect: The effect of presentation format. *Cognition, 175*, 141–156. https://doi.org/10.1016/j.cognition.2018.02.003

Cataldo, A. M., & Cohen, A. L. (2019). The comparison process as an account of variation in the attraction, compromise, and similarity effects. *Psychonomic Bulletin & Review, 26*(3), 934–942.

Cataldo, A. M., & Cohen, A. L. (2021). The influence of within-alternative and within-dimension similarity on context effects. *Decision, 8*(3), 202.

Evangelidis, I., & van Osselaer, S. M. (2019). Interattribute evaluation theory. *Journal of Experimental Psychology: General, 148*(10), 1733.

Frederick, S., Lee, L., & Baskin, E. (2014). The limits of attraction. *Journal of Marketing Research, 51*(4), 487–507.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research, 9*(1), 90–98.

Huber, J., Payne, J. W., & Puto, C. P. (2014). Let's be honest about the attraction effect. *Journal of Marketing Research, 51*(4), 520–525.

Katsimpokis, D., Fontanesi, L., & Rieskamp, J. (2022). A robust bayesian test for identifying context effects in multiattribute decision-making. *Psychonomic Bulletin & Review,* 1–18.

Noguchi, T., & Stewart, N. (2014). In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions. *Cognition, 132*(1), 44–56.

Noguchi, T., & Stewart, N. (2018). Multialternative decision by sampling: A model of decision making constrained by process data. *Psychological Review, 125*(4), 512.

Otto, A. R., Devine, S., Schulz, E., Bornstein, A. M., & Louie, K. (2022). Context-dependent choice and evaluation in real-world consumer behavior. *Scientific Reports, 12*(1), 17744.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychological Review, 108*(2), 370.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research, 16*(2), 158–174.

Simonson, I., Bettman, J. R., Kramer, T., & Payne, J. W. (2013). Comparison selection: An approach to the study of consumer judgment and choice. *Journal of Consumer Psychology, 23*(1), 137–149.

Spektor, M. S., Bhatia, S., & Gluth, S. (2021). The elusiveness of context effects in decision making. *Trends in Cognitive Sciences, 25*(10), 843–854.

Trendl, A., Stewart, N., & Mullett, T. L. (2021). A zero attraction effect in naturalistic choice. *Decision, 8*(1), 55.

Trueblood, J. S. (2015). Reference point effects in riskless choice without loss aversion. *Decision, 2*(1), 13–26.

Trueblood, J. S. (2022). Theories of context effects in multialternative, multiattribute choice. *Current Directions in Psychological Science, 31*(5), 428–435.

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological Review, 121*(2), 179.

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2015). The fragile nature of contextual preference reversals: Reply to Tsetsos, Chater, and Usher (2015). *Psychological Review, 122*(4), 848–853.

Trueblood, J. S., Liu, Y., Murrow, M., Hayes, W. M., & Holmes, W. R. (2022). *Attentional dynamics explain the elusive nature of context effects* [PsyArXiv preprint].

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review, 79*(4), 281.

Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science, 39*(10), 1179–1189.

Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review, 111*(3), 757.

Wu, C., & Cosguner, K. (2020). Profiting from the decoy effect: A case study of an online diamond retailer. *Marketing Science, 39*(5), 974–995.

Yang, S., & Lynn, M. (2014). More evidence challenging the robustness and usefulness of the attraction effect. *Journal of Marketing Research, 51*(4), 508–513.