Graphical Abstract

Unmasking the Flaws of Triplet-Triplet Attraction Effect Measures: Via Mathematical Analysis and Agent-Based Simulations

Tapas Rath, Nisheeth Srivastava, Narayanan Srinivasan

Highlights

Unmasking the Flaws of Triplet-Triplet Attraction Effect Measures: Via Mathematical Analysis and Agent-Based Simulations

Tapas Rath, Nisheeth Srivastava, Narayanan Srinivasan

- Identifies metric flaws causing false alarms and misses in attraction effect detection under preference bias. (Key finding: RST/ΔP show spurious effects; AST/ASC mask opposing decoy influences.)
- Recommends baseline comparisons (pair-triplet design) or baseline symmetry checks to mitigate measurement errors. (Key contribution: Methodological reforms for robust context effect detection.)

Unmasking the Flaws of Triplet-Triplet Attraction Effect Measures: Via Mathematical Analysis and Agent-Based Simulations

Tapas Rath[a], Nisheeth Srivastava[a], Narayanan Srinivasan[a]

[a] *Cognitive Science, Indian Institute of Technology Kanpur, ESB02, 4th Floor, Kanpur, 208016, UP, India*

Abstract

The 'attraction effect' or 'asymmetric dominance effect' is a widely studied phenomenon in decision-making, challenging the principle of regularity in rational choice theory. It posits that the introduction of a third option, similar but inferior to one of the available options in a binary choice set, increases the choice share of the dominating option. While the standard attraction effect is pervasive, recent studies have found inconsistent results. Some researchers attribute these inconsistencies to boundary conditions that constrain the effect, suggesting that studies failing to meet these criteria unsurprisingly report inconsistent effects. This paper aims to clarify the presence, strength, and direction of the attraction effect through two approaches: first, by analyzing contemporary metrics for measuring context effects and demonstrating their susceptibility to false positives and negatives; and second, by examining the boundary condition of strong prior trade-off—a biased preference for one choice in the baseline set—and its impact on metric vulnerability.

Keywords: Triplet-triplet design, Attraction effect, Simulation

1. Introduction

According to one of the classic principles of rational decision-making —*regularity* (Luce, 1977)— if a choice set C_1 is a subset of a larger choice set C_2, then the probability of choosing any option x from C_2 should not exceed its probability of being chosen from C_1. Formally:

$$\forall x \in C_1 \subseteq C_2, \quad P(x \mid C_1) \geq P(x \mid C_2).$$

Intuitively, adding an alternative to a choice set should not increase the choice probability of any existing alternative.

As an empirical challenge to this principle, Huber et al. (1982) demonstrated the now-famous *attraction effect* (or *asymmetric dominance effect*), where the presence of a decoy alternative increases the choice share of a particular option. Over the past four decades, numerous pure, conceptual, and domain replications of the attraction effect have been reported. However, more recent studies have documented null or even reversed effects, leading some researchers to describe the attraction effect as "fragile" or "elusive" (Spektor et al., 2021) and to question its "practical validity" (Frederick et al., 2014). In response to these concerns, Huber et al. (2014) outlined various *boundary conditions* for the effect and recommended manipulation checks in future studies.

This paper focuses on one such boundary condition: *prior trade-offs* in the core set and the factors that generate them. If the choice probabilities in the core set deviate significantly from an even (e.g., 50–50) split, this indicates a strong pre-existing trade-off that may be *resistant* to decoy manipulations. **?** suggested several factors underlying prior trade-offs, including individual differences, attribute importance ratings, and practice effects. Further, several other researchers (Trueblood, 2015; Simonson, 2014; Katsimpokis et al., 2022) have also raised concerns about *heterogeneity in values across individuals*, and how strong attribute preferences or dimensional biases (Hutchinson et al., 2000; Liew et al., 2016) might *attenuate or reverse* the measured effect sizes. Notably, **?** listed both *prior trade-offs* and *cross-respondent value heterogeneity* as drivers that can inhibit attraction effects. However, to date, there has been *no rigorous mathematical analysis* of how such prior bias influences attraction effects. We raise a slightly different point in this paper: we suggest that there are measurement errors, and that these errors are exacerbated by the prior bias.

To support our claims, we first analyze contemporary measures used for context effects, including RST, RST_{ew}, and AST, demonstrating their analytical vulnerabilities. We then employ an agent-based modeling framework, simulating decision-makers with heterogeneous subjective indifference curves (SICs), to further illustrate how prior bias leads to false alarms and misses when using these metrics in triplet-triplet designs.

The remainder of this paper is structured as follows. Section 2 provides an overview of context effects, including the similarity and attraction effects. Section 3 details the measures of context effects and experimental designs, focusing on pair-triplet and triplet-triplet designs and metrics like ΔP_{target}, RST, RST_{ew}, AST, and ASC. Section 4 presents analytical assessments and simulations demonstrating the vulnerability of RST, ΔP, and AST to false alarms and misses, particularly exacerbated by prior bias. Section 5 introduces hypothetical choice models and agent-based simulations to further illustrate these measurement failures at the individual level. Section 6 discusses the implications of our findings for the current controversy surrounding the attraction effect. Finally, Section 7 concludes by summarizing our findings and proposing potential solutions.

2. Context Effects

The major context effects discussed in the literature are the *similarity effect* (Tversky, 1972), the *attraction effect* (Huber et al., 1982), and the *compromise effect* (Simonson, 1989). In this article, however, we focus primarily on the attraction effect, using the similarity effect as additional empirical evidence within one of our hypothetical choice model simulations.

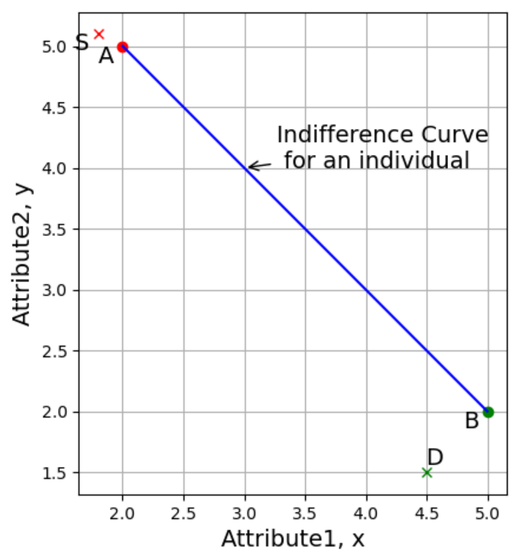


Figure 1: Illustration of the Similarity Effect and Attraction Effect.

Notes: S is the similarity decoy targeting B, and D is the attraction decoy targeting B. The x- and y-axes represent attributes 1 and 2, respectively.

If an item S is added to a binary choice set $\{A, B\}$ such that items A and S are placed close to each other in attribute space (as shown in Figure 1) and are perceived as similar, the choice share of B increases:

$$P(B|\{A, B, S\}) > P(B|\{A, B\}).$$

This is known as the *similarity effect* (Tversky, 1972), which, in some form, predicted violations of *stochastic transitivity*—a weaker form of the rational choice rule known as *Independence from Irrelevant Alternatives* (IIA).[1] However, the model Tversky proposed still adhered to the regularity principle. Later, Luce (1977) argued that regularity was the only rational choice axiom that had remained unviolated.

As discussed earlier, the attraction effect demonstrates a violation of regularity. Suppose items A and B lie on the same indifference curve in attribute space (as depicted in Figure 1), and a third item D—inferior to B—is added to the set. If a clear dominance relationship is perceived between B and D (such that D is dominated by B but not by A), then the choice share of B increases when moving from the binary to the ternary set:

$$P(B|\{A, B, D\}) > P(B|\{A, B\}).$$

3. Metrics for Decoy Effects

The key rational choice assumptions of Independence of Irrelevant Alternatives (IIA) and regularity were introduced in previous sections. To examine potential violations of these principles, we define several metrics that capture the biased choice behavior.

3.1. Target Probability Difference (ΔP_{target})

This metric captures the change in the target's choice probability between the triplet context (where a decoy is present) and the pair context (without the decoy):

$$\Delta P_{target} = P(\text{Target}_{C1}) - P(\text{Target}_{C0}).$$

In choice frequencies:

$$\Delta P_{target} = \frac{n_{\text{Target},C1}}{n_{\text{Target},C1} + n_{\text{Competitor},C1} + n_{Decoy,C1}} - \frac{n_{\text{Target},C0}}{n_{\text{Target},C0} + n_{\text{Competitor},C0}}.$$

[1] Strong IIA requires that the ratio of choice shares between any two items remains constant, regardless of the choice set in which they are presented.

Interpretation:

- $\Delta P_{target} > 0$ indicates a violation of the regularity assumption, as the target's choice probability increased when the decoy was added.
- $\Delta P_{target} \leq 0$ indicates perfect consistency with regularity.

Note that ΔP_{target} is expected to have a reference value of zero.

3.2. ΔP_{target} in Triplet-Triplet Design

To enhance effect size and hence the statistical power of studies, Wedell (1991) introduced a triplet-triplet design, where the focal options A and B are each presented with a decoy that favors one option in one context and the other option in the second context—that is, $\{A, B, D_a\}$ and $\{A, B, D_b\}$, where D_a favors A and D_b favors B. Wedell claimed that this design offers a double opportunity to detect context effects compared to the simpler pair-triplet design.

Subsequently, most context effect studies have adopted this triplet-triplet design. One measure of the effect in such a design, commonly used by Wedell (1991) and others (Liu and Trueblood, 2023), is to compute the difference in the target's choice share between the two contexts:

$$\Delta P_{target} = P(A \mid \{A, B, D_a\}) - P(A \mid \{A, B, D_b\}).$$

Here, $P(A \mid \{A, B, D_x\})$ represents the probability of choosing option A in the presence of decoy D_x.

Interpretation:

- $\Delta P_{target} > 0$ indicates an attraction effect (the decoy favoring A increases A's choice probability relative to the context where the decoy favors B).
- $\Delta P_{target} < 0$ indicates a reversed attraction effect (the decoy that supposedly favors A actually reduces A's choice probability relative to the other context).
- $\Delta P_{target} = 0$ indicates no differential decoy effect between contexts (consistent with regularity and IIA).

3.3. Relative Share of Target (RST)

An alternative to ΔP_{target} is the *Relative Share of Target* (RST), proposed by Berkowitsch et al. (2014).

In terms of choice frequencies, aggregating target choices across both contexts and normalizing by the total choices of the target and competitor:

$$RST = \frac{n_{\text{Target},C1} + n_{\text{Target},C2}}{n_{\text{Target},C1} + n_{\text{Target},C2} + n_{\text{Competitor},C1} + n_{\text{Competitor},C2}}.$$

An RST value above 0.5 indicates that the target is chosen more often than the competitor, while an RST of 0.5 indicates equal preference.

3.4. Equal-Weight Relative Share of the Target (RST_{ew})

To reduce potential biases in RST, Katsimpokis et al. (2022) proposed RST_{ew} (equal weights), which averages the target share within each context:

$$RST_{ew} = \frac{1}{2}\left(\frac{P(\text{Target}_{C1})}{P(\text{Target}_{C1}) + P(\text{Competitor}_{C1})} + \frac{P(\text{Target}_{C2})}{P(\text{Target}_{C2}) + P(\text{Competitor}_{C2})}\right).$$

The corresponding choice frequency formula is:

$$RST_{ew} = \frac{1}{2}\left(\frac{n_{\text{Target},C1}}{n_{\text{Target},C1} + n_{\text{Competitor},C1}} + \frac{n_{\text{Target},C2}}{n_{\text{Target},C2} + n_{\text{Competitor},C2}}\right).$$

Values above 0.5 indicate a higher relative preference for the target on average, while values below 0.5 indicate a higher preference for the competitor.

3.5. Absolute Shares of the Target and Competitor (AST and ASC)

While the triplet-triplet design primarily tests for *menu dependence* (violation of IIA), it can also be used to indirectly test for violations of the *regularity principle* (Katsimpokis et al., 2022). Katsimpokis et al. (2022) introduced the *Absolute Share of Target* (AST) and the *Absolute Share of Competitor* (ASC), which incorporate decoy choices into the denominator:

Absolute Share of the Target (AST)..

$$AST = \frac{1}{2}\left(P(\text{Target}_{C1}) + P(\text{Target}_{C2})\right).$$

In choice frequencies:

$$AST = \frac{1}{2}\left(\frac{n_{\text{Target},C1}}{n_{\text{Target},C1} + n_{\text{Competitor},C1} + n_{D,C1}} + \frac{n_{\text{Target},C2}}{n_{\text{Target},C2} + n_{\text{Competitor},C2} + n_{D,C2}}\right).$$

Absolute Share of the Competitor (ASC)..

$$ASC = \frac{1}{2} \left(P(\text{Competitor}_{C1}) + P(\text{Competitor}_{C2}) \right).$$

In choice frequencies:

$$ASC = \frac{1}{2} \left(\frac{n_{\text{Competitor},C1}}{n_{\text{Target},C1} + n_{\text{Competitor},C1} + n_{D,C1}} + \frac{n_{\text{Competitor},C2}}{n_{\text{Target},C2} + n_{\text{Competitor},C2} + n_{D,C2}} \right).$$

An AST value greater than 0.5 indicates a standard (positive) attraction effect favoring the target, while an ASC value greater than 0.5 indicates a reversed (negative) attraction effect favoring the competitor.

Note that AST and ASC are adapted from Katsimpokis et al. (2022), and their derivation is provided in Appendix A. while ΔP_{target} has a reference value of zero, the triplet-triplet metrics (RST, RST_{ew}, AST, and ASC) have a reference value of 0.5, corresponding to equal preference between the target and competitor. To facilitate comparison across metrics, these can be converted to a zero-referenced format by subtracting 0.5, with positive values in RST, RST_{ew}, AST indicating relative preference for the target.

4. Analytical Proofs of Metric Vulnerabilities

4.1. RST False Positives Under IIA

In probability terms,

$$\text{RST} = \frac{P_{C_1}(A) + P_{C_2}(B)}{P_{C_1}(A) + P_{C_2}(B) + P_{C_1}(B) + P_{C_2}(A)} \tag{1}$$

Substituting:

$$\text{RST} = \frac{rb_1 + b_2}{rb_1 + b_2 + b_1 + rb_2} \tag{2}$$

where $b_1 = P_{C_1}(B)$, $b_2 = P_{C_2}(B)$, and $r = \frac{P_{C_1}(A)}{P_{C_1}(B)} = \frac{P_{C_2}(A)}{P_{C_2}(B)}$ (Assuming strong IIA holds).

4.1.1. Condition for RST = 0.5

$$\frac{rb_1 + b_2}{rb_1 + b_2 + b_1 + rb_2} = \frac{1}{2} \tag{3}$$

$$2(rb_1 + b_2) = rb_1 + b_2 + b_1 + rb_2 \tag{4}$$

$$\Rightarrow (rb_1 - b_1) = (rb_2 - b_2) \tag{5}$$

$$\Rightarrow b_1(r - 1) = b_2(r - 1) \tag{6}$$

Thus,

$$\text{RST} = 0.5 \quad \text{if and only if} \quad r = 1 \quad \text{or} \quad b_1 = b_2 \tag{7}$$

Implication: When the preference ratio $r \neq 1$ (i.e., prior bias exists), RST will deviate from 0.5 even if the underlying preference structure is unchanged, leading to *false positives*.

4.2. Misses Under Opposing Effects

Let the baseline choice probabilities between options A and B be:

$$P_0(A) = p, \quad P_0(B) = 1 - p, \tag{8}$$

where $p \in (0, 1)$.
Let us define two triplet contexts: Context 1 (C_1), defined as $\{A, B, D_A\}$, with decoy D_A designed to favor A (target: A); and Context 2 (C_2), defined as $\{A, B, D_B\}$, with decoy D_B designed to favor B (target: B). The choice probabilities are:

$$P(\text{Target}_{C1}) = P_{C_1}(A) = p + \Delta_A, \quad P_{C_1}(B) = (1 - p) - \Delta_A - \delta_1, \tag{9}$$

$$P(\text{Target}_{C2}) = P_{C_2}(B) = (1 - p) + \Delta_B, \quad P_{C_2}(A) = p - \Delta_B - \delta_2. \tag{10}$$

where Δ_A and Δ_B are the effect sizes in contexts C_1 and C_2 respectively, and δ_1 and δ_2 are the probabilities of the decoys in the respective contexts. As introduced earlier,

$$AST = \frac{1}{2}\left(P(\text{Target}_{C1}) + P(\text{Target}_{C2})\right).$$

From the equations:

$$P(\text{Target}_{C1}) = p + \Delta_A \quad \text{and} \quad P(\text{Target}_{C2}) = (1 - p) + \Delta_B.$$

Thus:

$$AST = \frac{1}{2}\left[(p + \Delta_A) + ((1 - p) + \Delta_B)\right] = \frac{1}{2}\left[1 + (\Delta_A + \Delta_B)\right].$$

Subtracting 0.5:

$$AST - 0.5 = \frac{\Delta_A + \Delta_B}{2}.$$

When decoys produce opposing effects ($\Delta_A > 0, \Delta_B < 0$):

$$AST - 0.5 = \frac{\Delta_A + \Delta_B}{2} \leq \frac{\Delta_A}{2} < \Delta P_A.$$

AST underestimates true effects by at least 50%. A similar analysis can be applied to all other triplet-triplet metrics that use averaging across two contexts, demonstrating that such metrics systematically underestimate the effect sizes when the two decoys produce opposing effects.

5. Simulation Studies

5.1. Parameter Space Exploration

5.1.1. For RST metric

To assess the sensitivity of the RST metric to variations in baseline preferences and decoy probabilities, we conducted a systematic parameter space exploration. Specifically, we examined the behavior of RST as defined in Equation (**??**), which depends on the baseline preference ratio r and the baseline choice probabilities b_1 and b_2 associated with the decoy-influenced choices in the two contexts.

In our simulation, the baseline preference ratio r was varied continuously from 0.5 to 2.0, while b_1 and b_2 were varied independently within the range 0.05 to 0.45. For each combination of r, b_1, and b_2, we computed the corresponding RST value. We then identified, for each r, the minimum and maximum RST values across the full range of b_1 and b_2.

The results of this analysis revealed that RST systematically deviates from the null value of 0.5 whenever either $r \neq 1$ or $b_1 \neq b_2$, even in the absence of any true context effect. This finding demonstrates that RST can produce apparent effects purely due to baseline asymmetries, thereby generating false positives. These results underscore the need to interpret RST with caution, especially when baseline preference asymmetries are present. Figure 2 shows the deviations of RST from 0.5.

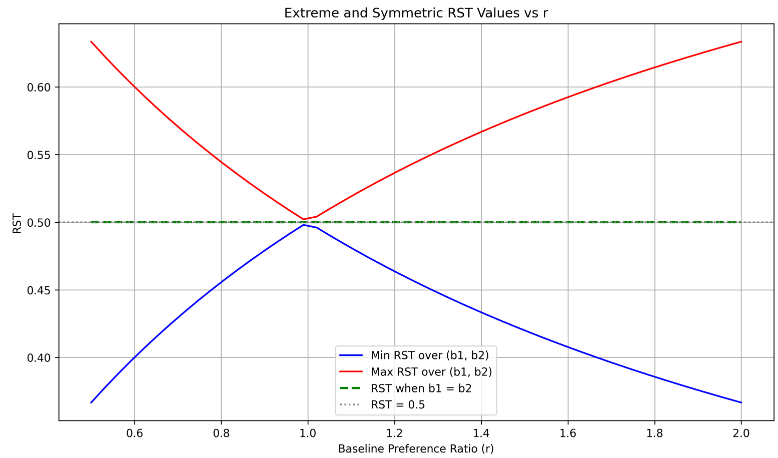


Figure 2: RST values across different baseline preference ratios r. Deviations from 0.5 occur when $r \neq 1$, indicating false alarms.

5.1.2. For AST metric

We next examined the behavior of the AST metric, recently introduced as a triplet-triplet measure of context effects, under varying decoy-induced effects across two different contexts. As noted earlier, triplet-triplet measures—including RST, RST_{ew}, and AST or ASC—rely on the critical assumption that the decoys introduced in these two contexts behave symmetrically relative to participants' subjective valuations. This assumption is often operationalized by ensuring equal target–decoy distances in both contexts to claim decoy symmetry.

However, as discussed previously, baseline biases—such as a prior trade-off in favor of one attribute dimension or an inherent preference for one of the core items (A or B)—violate this assumption of symmetry. When SICs shift due to such preferences, the decoys no longer exert equivalent influences in both contexts. In this scenario, a decoy in one context may produce a positive attraction effect, while the decoy in the other context may produce a negative (or reversed) attraction effect. When averaged across contexts, the metrics can fail to detect these opposing effects, leading to an underestimation or even complete cancellation of the true context effect. Consequently, these metrics may miss violations of regularity and the Independence of Irrelevant Alternatives (IIA) between each triplet and the baseline pair, while potentially indicating no violation between the two triplet contexts themselves.

To illustrate this, we simulated the AST surface by varying Δ_A and Δ_B

independently from -0.2 to 0.2. Here, Δ_A and Δ_B represent the decoy-induced context effects in each of the two contexts, respectively. For each combination, we calculated the AST value using the relation:

$$\mathrm{AST} = 0.5\big(1 + \Delta_A + \Delta_B\big),$$

and plotted the deviation from the null value of 0.5 in a heatmap. The results, shown in Figure X, reveal that when $\Delta_A > 0$ and $\Delta_B < 0$, AST approaches the null value (< 0.5) despite substantial positive and negative context effects in the individual triplets. This cancellation is particularly evident along the diagonal $\Delta_A = -\Delta_B$, where the patterned region below this line indicates AST < 0.5. This highlights that the standard attraction effect is effectively masked, leading to false negatives even when genuine context effects are present. Consequently, AST is insensitive to opposing decoy effects in the two contexts and may miss violations of regularity and IIA. This analysis underscores the limitations of AST in detecting context effects when decoys exert asymmetric or opposing influences across the two contexts.

5.2. Agent-Based Modeling

In the following section, we adopt hypothetical choice models and further simulate choices under different SICs to illustrate the limitations of triplet-triplet metrics. We use an agent-based modeling framework in which agents (decision-makers) are simulated independently, each characterized by a unique subjective indifference curve. No agent–agent interaction is modeled; our focus is on heterogeneity in individual choice behavior and its aggregate implications.

Our first set of agents (hypothetical choice models) mimic ideal (rational) decision-makers who perfectly obey IIA. Nonetheless, we show how most of the measures produce false alarms. In contrast, our second hypothetical choice model uses empirical evidence to simulate decision-makers who violate regularity and hence IIA between pair and triplet contexts. In this case, we demonstrate how the metrics fail to detect these violations and instead falsely report null effects or reduced effects.

5.2.1. Simulation Methodology

We used a Python-based program to simulate agents with different SICs. Items in the choice set were represented as points on the x-y plane. The

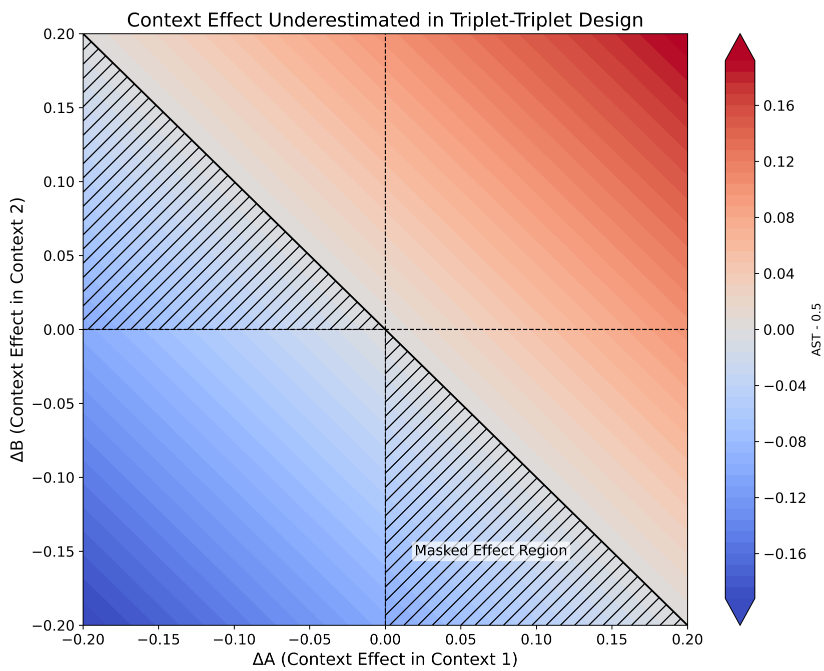


Figure 3: Heatmap of $AST - 0.5$ over Δ_A and Δ_B. Patterned regions indicate underestimation or complete masking of effects.

x-y plane represents the attribute space, with the two primary attributes of interest representing the x and y-axis. Two items, A and B, from the binary choice set are placed on the attribute space as two points, and we call the line joining them, i.e., AB, the experimenter-defined indifference curve (EDIC). Note, here, that we assume a linear indifference curve for simplicity. True SICs representing different individuals are lines with varying slopes. We assume that slope alone can characterize a linear indifference curve, meaning an infinite number of parallel indifference curves on this plane can represent one specific individual. Leveraging this property, we simulated all SICs as lines with different slopes, passing through point A, and proceeded with further mathematical analysis. Suppose two points U and W lie on the indifference curve. In that case, it means the corresponding individual is indifferent to the two items represented by the points, i.e., when the subject comes across the items in binary choice sets appearing enough times in between other trials, his choice frequencies for both will be roughly equal, i.e. $P(U|\{U, W\}) \approx P(W|\{U, W\}$.

A deviation from the SIC is characterized by a signed perpendicular distance from it, such that points right to the SIC will have a positive deviation and points to the left will have a negative deviation. As the direction along the positive x-axis represents the increasing strength of the attribute, items with positive deviation are considered 'better' than the corresponding foot of the perpendicular on the SIC. This assumption is reasonable. The preference accumulation model by Bhatia (2013) also has a similar assumption.

The signed deviations from the SIC serve as arguments to a softmax function [2] that outputs choice shares in terms of probabilities summing to one. Figure 4a shows line AB as the EDIC with the perpendicular line segments BB' representing the signed deviation of B from the B' points on respective SICs. Figure 4b shows the simulated baseline choice shares in the binary choice set as a function of SIC slopes.

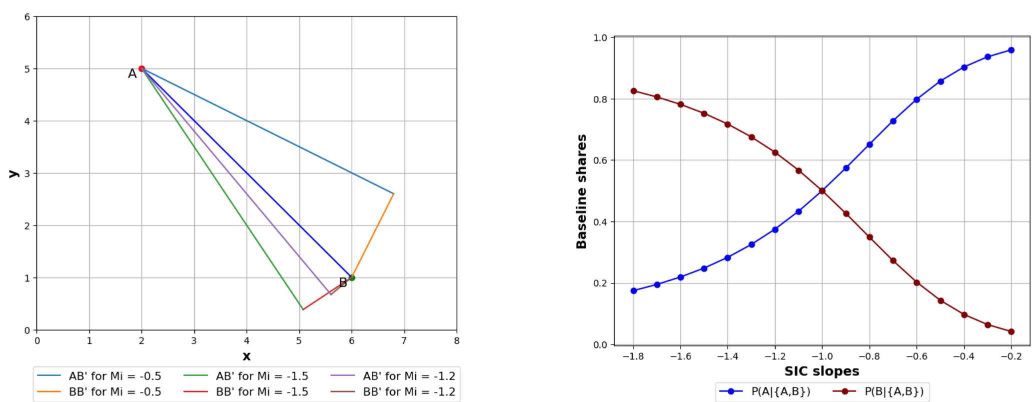
We used a Temperature-Scaled SoftMax function to model choice shares. As mentioned earlier, a vector of signed perpendiculars corresponding to each element in the choice set serves as the argument for the softmax.

$$\text{softmax}(x_i) = \frac{e^{(\beta \cdot (x_i - \max(x)))}}{\sum_{j=1}^{N} e^{(\beta \cdot (x_j - \max(x)))}},$$

[2] A justification of using a softmax function is discussed in the Appendix.

where, x_i represents the i-th element of the input vector. $\max(x)$ calculates the maximum value within the input vector x. Subtracting the maximum value $\max(x)$ from the input values provides numerical stability. It ensures that the largest exponent in the numerator is zero, preventing overflow issues that might occur when dealing with large numbers in the exponential function. β is a multiplicative coefficient that scales the input values before computing the softmax function, i.e., a higher β amplifies the differences between the input values. N denotes the size of the choice set: 2 for the pair and 3 for the triplets. The softmax function normalizes the scaled exponentiated values by dividing each exponentiated input by the sum of all exponentiated inputs. This normalization ensures that the output values lie in the range [0, 1] and sum up to 1, representing a valid probability distribution over the input values.

Figure 4: Subject Specific Indifference Curves (SICs)



(a) SICs represented by different slopes M_i

(b) Baseline Shares vs SIC Slopes

Figure 4a shows line AB as the EDIC with the perpendicular line segments BB' representing the signed deviation of B from the B' points on respective SICs. Figure 4b shows the simulated baseline choice shares in the binary choice set as a function of SIC slopes.

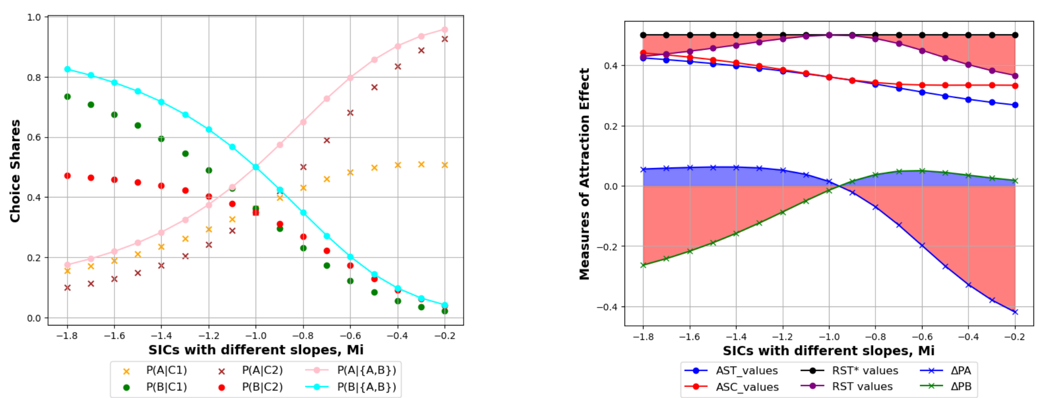
As we describe below, temperature scaling offers an effective parametric way of instantiating theoretical expectations for context effects. In our simulations, items A and B are presented either in pairs or in two triplets $\{A, B, D_a\}$ or $\{A, B, D_b\}$, where D_a and D_b are two decoy options targeting A and B, respectively. The two triplets serve as two choice contexts.

5.2.2. Model 1 (IIA-Adherent)

In the absence of any context effect, Luce's choice rule, independence from irrelevant alternative (IIA) should be adhered to, i.e., the ratio of choice shares of two items should remain constant irrespective of the context (choice set) in which they are presented; $\frac{P(X|C)}{P(Y|C)} =$ Constant, irrespective of C; where X and Y are two items and C is any context. This condition is easily achieved by setting $\beta = 1$ in the softmax function.

Results. Figure 5b clearly shows that some of the measures of context effects (RST and ΔP) show negative effects for nearly all values of SIC slopes. We call these false alarms (FAs) in the detection of context effects. Notably, RST_{ew}, AST, and ASC do not appear to demonstrate false alarms, unlike other measures.

Figure 5: False Alarms in the Detection of Context Effects



(a) Choice Shares

(b) Measures of Context Effects

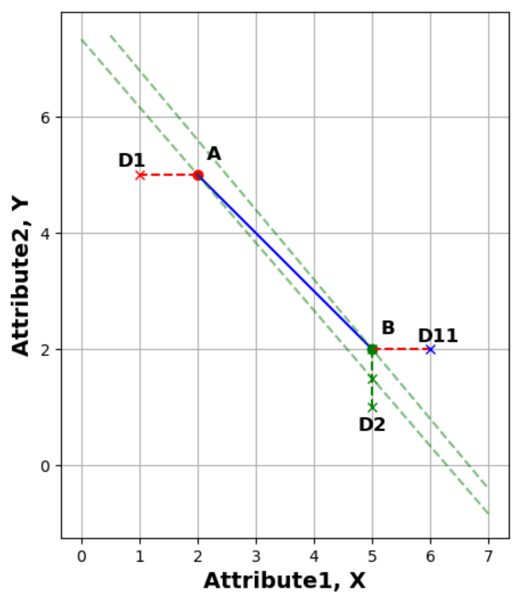
Notes: C: $C1(\{A, B, D_a\})$, $C2(\{A, B, D_b\})$. The choice shares of A and B in the triplets have values less than their corresponding baseline values in plot 5a. In 5b, the standard attraction effect is shown in a blue-filled color, and the reversed attraction effect is shown in a red-filled color.

5.2.3. Model 2 (IIA-Violating)

Our modeling goal, here, was to artificially introduce a violation of IIA, mimicking one of the context effects as observed from the empirical data, such that in one of the two triplets, there is an attraction effect, while in the other, there is a reverse attraction effect. This would allow us to test the efficacy of the instruments used to capture context effects in a triplet-triplet design. We consider two conditions:

1. When there is a dominance perceived between the target and the attraction decoy, the choice share of the target increases in the triplet. (Attraction effect)
2. When a dominance relation is not perceived between the target and the decoy; rather, both are perceived to be similar, the choice share of the competitor would increase. (A similarity-effect driven reverse-attraction effect)

Figure 6: Attraction and reversed attraction in the two contexts



As shown in Figure 6, in this simulation, the first condition is achieved when an SIC passing through A also passes through the line segment joining B and D2. This would make the points B and D2 fall on two opposite sides of the SIC, hence assigning opposite signs to the corresponding signed perpendiculars that will serve as the arguments to the softmax function. In other words, B is perceived as 'better' than its corresponding foot of the perpendicular (and hence point A) on the SIC, while D2 is perceived as 'worse' than A. This makes a clear perceived dominance relationship between B and D2, satisfying the condition for the attraction effect (the first theory considered

in this simulation). We model the effect on choice shares by simply setting $\beta = 4.5$ in the softmax function. This computation increases the choice share of B, reasonably mimicking an attraction effect targeting B.
Note, however, that for the same SIC, points A and D1 lie on the same side of the curve. This can be appreciated by noticing the parallel line passing through point B, but having the same slope, hence representing the same SIC. Because of this, both A and D1 are perceived to be worse as compared to B. We model this as a case of similarity where a clear dominance between A and D1 is not perceived; rather, both appear similar with respect to B. Hence, the similarity effect should operate, and we achieve it by again setting $\beta = 4.5$, which in turn, increases the choice share of B. This is technically a similarity-induced reversed attraction effect.
Taken together, for SIC slopes in a small range close to EDIC, for the same SICs, when D1s produce reversed attraction effects, D2s produce attraction effects. Similarly, it can be shown that for an SIC through A, that passes between B and D11 (D11 is simply a proxy for D1 when the SIC is shown to have passed through A rather than through B), D2 produces a reversed attraction effect while D1 produces an attraction effect.

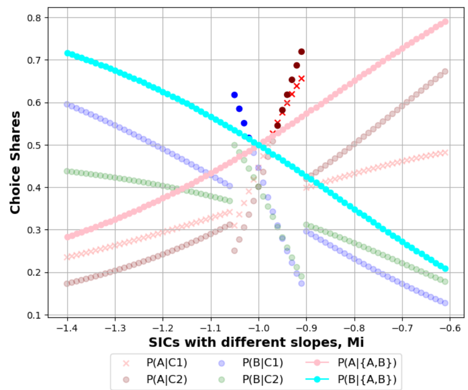
Results. Figure 7 shows results from our simulations for an individual agent. In particular, the filled dots in Figure 7a indicate instances of choice shares of A and B in triplets (C1: $\{A, B, D_a\}$, C2: $\{A, B, D_b\}$) whose values are clearly beyond their corresponding baseline values. These are cases of violations of regularity, a weaker form of IIA. Figure 7b shows how no existing measures of context effects, including the recently introduced AST and ASC, capture them. We call such instances 'misses' in the context of detecting context effects.
Notes: C: $C1(\{A, B, D_a\})$, $C2(\{A, B, D_b\})$. 7b shows only reversed attraction effects as indicated by RST.

6. Discussion

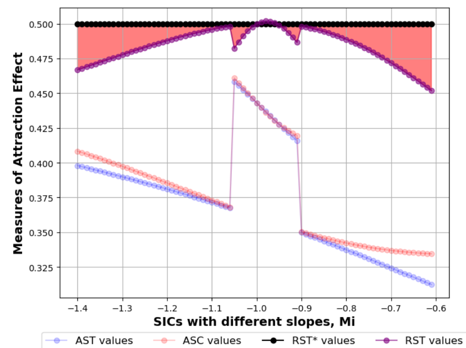
Our combined analysis reveals critical limitations:

- **False alarms**: RST and ΔP_{Target} signals IIA violations due solely to baseline asymmetries in core choice probabilities, not genuine effects.
- **Missed effects**: AST, ASC, and RST_{ew} cancel opposing effects, leading to underestimation or complete masking of true context effects.

Figure 7: Misses in Detecting Context Effects



(a) Choice Shares



(b) Measures of Context Effect

It is conceivable that the measurement error in the presence of dimensional bias, which we discussed in the paper, is one of the key reasons for the empirical controversy surrounding the characterization of the attraction effect (Huber et al., 2014; Frederick et al., 2014; Spektor et al., 2018, 2022).For example, in the 12 experiments, 1A-1S in (Frederick et al., 2014), all the studies have biased baseline choice shares, and they found mostly reversed effects. Similarly, Liew et al. (2016) have shown dimensional biases, but have concluded that averaging across the population consisting of dimensional biases is the cause of reduced context effects. While that is a valid conclusion, we make a different case here; two contexts may have different effects for the same individuals owing to the asymmetry of decoy placements with respect to the SICs. It should be noted that following Wedell (1991), most studies showing reversed or reduced attraction effects (Spektor et al., 2018, 2022) have employed triplet-triplet designs, and as they have not reported binary baseline choice shares, the confounds we have highlighted in this paper apply to their results as well. We argue that for testing context-based preference reversals, either the pair-triplet design should be incorporated, or if the triplet-triplet design is employed, experimenters need to separately ensure that there is symmetry of choice shares between the baseline core options. One of the limitations of our work is, as mentioned earlier, we have just focused on one of the boundary conditions of the effect that Huber et al. (2014) have already discussed, i.e., ‘strong prior trade-off’ and showed how the triplet-triplet measure could be misleading in the presence of the biased baseline shares. So, our work here adds to theirs. Also, we acknowledge that we have used only range decoys for the agent-based simulations, as they are known to produce maximum context effects. However, the simulations can be easily extended to accommodate other decoy types without affecting the interpretations.

7. Conclusion

Existing triplet-based metrics are fundamentally flawed: they are prone to both false positives and false negatives due to their structural assumptions. As solutions to the issues with the extant instruments, we propose including pair-triplet measurements, wherever possible, in the experimental designs and accessing the subject-specific indifference curves to design the stimulus space accordingly. With the growing influence of digitization and user profiling in online shopping, baseline choice shares have become more accessible for marketers to leverage. Hence, the issues discussed and the proposed solutions equally apply to marketers who wish to incorporate the context effects into product designs and marketing strategies.

Appendix A. Justification of SoftMax as a model of choice of shares

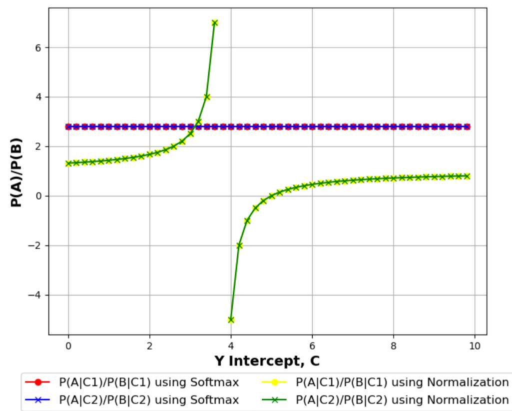
We could have used a simple normalization instead of using SoftMax as the model of choice in the simulations. Both normalization and SoftMax maintain the ratio of choice shares between the two contexts without violating strong IIA. Note in Figure A.8 that

$$\frac{P(A \mid C_1)}{P(B \mid C_1)} = \frac{P(A \mid C_2)}{P(B \mid C_2)}$$

for both the models. However, the assumption that all parallel lines having the same slopes represent one SIC can only be supported by the SoftMax model, as the ratio of choice probabilities it produces remains constant over varying y intercepts of the SIC for a specific slope. Note that this is clearly not the case with normalization, as shown in the figure below.

Notes: The plot shows ratios of choice shares of A and B as a function of different y-intercepts of SIC having a specific slope, m = -0.6.

Justification of SoftMax over normalization as a model of choice.



Appendix B. Derivations of AST and ASC

The derivation is adapted from (Katsimpokis et al., 2022). Suppose A and B are two core options presented in the binary choice set $\{A, B\}$ and in ternary choice sets $\{A, B, D_a\}$ and $\{A, B, D_b\}$, where D_a and D_b are the decoys favoring A and B respectively in the two contexts. Adhering to the principle of regularity,

$$P(A|\{A,B\}) \geq P(A|\{A,B,D_a\}), \tag{1}$$
$$P(A|\{A,B\}) \geq P(A|\{A,B,D_b\}), \tag{2}$$
$$P(B|\{A,B\}) \geq P(B|\{A,B,D_a\}), \tag{3}$$
$$P(B|\{A,B\}) \geq P(B|\{A,B,D_b\}). \tag{4}$$

But, we know that according to the law of total probability, $P(A|\{A,B\}) + P(B|\{A,B\}) = 1$. So, adding equations 1 and 4,

$$P(A|\{A,B,D_a\}) + P(B|\{A,B,D_b\}) \leq 1. \tag{5}$$

Similarly, adding equations 2 and 3,

$$P(A|\{A,B,D_b\}) + P(B|\{A,B,D_a\}) \leq 1. \tag{6}$$

(Note that the other combinations would give rise to obvious results owing to the definition of multinomial distribution; they are not mentioned, and equations 5 and 6 can serve as tests for regularity violation.) To keep the format of the inequalities similar to that of RST, an existing measure of context effect (Katsimpokis et al., 2022) both sides of the above equations multiplied by 0.5, leading to

$$0.5 \times [P(A|\{A, B, D_a\}) + P(B|\{A, B, D_b\})] \leq 0.5, \tag{7}$$

$$0.5 \times [P(A'|\{A, B, D_b\}) + P(B|\{A, B, D_a\})] \leq 0.5. \tag{8}$$

Considering one of A and B as the target (t) and the other competitor (c), and the third as a decoy (D_{target}) favoring the target option, the probability terms in equations 5 and 6 can be rewritten in terms of choice frequencies as follows,

$$0.5 \times \left[\frac{n_A}{n_A + n_B + n_{D_a}} \text{ in } C_1 + \frac{n_B}{n_A + n_B + n_{D_b}} \text{ in } C_2 \right] \leq 0.5, \tag{9}$$

$$0.5 \times \left[\frac{n_B}{n_A + n_B + n_{D_a}} \text{ in } C_1 + \frac{n_A}{n_A + n_B + n_{D_b}} \text{ in } C_2 \right] \leq 0.5. \tag{10}$$

Where, n_x in $C\#$ is the choice frequency of option x in the context $C\#$; x is either A, B, D_a, or D_b. C_1 and C_2 are the two ternary choice sets $\{A, B, D_a\}$ and $\{A, B, D_b\}$, i.e., A is the target in C_1 and B is the target in C_2. The choice frequency of an option in a ternary choice set is equal to the total number of times that option is chosen when the ternary choice set appears enough times in between other trials containing other choice sets. Note that the two inequalities' left-hand sides are AST and ASC, respectively, as defined in the main text. AST and ASC should be less than or equal to 0.5 for no regularity violation.

References

Berkowitsch, N.A., Scheibehenne, B., Rieskamp, J., 2014. Rigorously testing multialternative decision field theory against random utility models. Journal of Experimental Psychology: General 143, 1331.

Bhatia, S., 2013. Associations and the accumulation of preference. Psychological review 120, 522.

Frederick, S., Lee, L., Baskin, E., 2014. The limits of attraction. Journal of Marketing Research 51, 487–507. doi:10.1509/jmr.12.0061.

Huber, J., Payne, J.W., Puto, C., 1982. Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis .

Huber, J., Payne, J.W., Puto, C.P., 2014. Let's be honest about the attraction effect. Journal of Marketing Research 51, 520–525. doi:10.1509/jmr.14.0208.

Hutchinson, J.W., Kamakura, W.A., Lynch, J.G., 2000. Unobserved heterogeneity as an alternative explanation for "reversal" effects in behavioral research. Journal of Consumer Research 27, 324–344. doi:10.1086/317588.

Katsimpokis, D., Fontanesi, L., Rieskamp, J., 2022. A robust bayesian test for identifying context effects in multiattribute decision-making. Psychonomic Bulletin Review doi:10.3758/s13423-022-02157-2. dOI: 10.3758/s13423-022-02157-2 MAG ID: 4297338249 PMID: 36167914.

Liew, S.X., Howe, P.D.L., Little, D.R., 2016. The appropriacy of averaging in the study of context effects. Psychonomic Bulletin Review 23, 1639–1646. doi:10.3758/s13423-016-1032-7.

Liu, Y., Trueblood, J.S., 2023. The effect of preference learning on context effects in multi-alternative, multi-attribute choice. Cognition 233, 105365. doi:10.1016/j.cognition.2022.105365.

Luce, R.D., 1977. The choice axiom after twenty years. Journal of Mathematical Psychology 15, 215–233. doi:10.1016/0022-2496(77)90032-3.

Simonson, I., 1989. Choice based on reasons: The case of attraction and compromise effects. Journal of Consumer Research 16, 158. doi:10.1086/209205.

Simonson, I., 2014. Vices and virtues of misguided replications: The case of asymmetric dominance. Journal of Marketing Research 51, 514–519. doi:10.1509/jmr.14.0093.

Spektor, M.S., Bhatia, S., Gluth, S., 2021. The elusiveness of context effects in decision making. Trends in Cognitive Sciences 25, 843–854. doi:10.1016/j.tics.2021.07.011. 24 citations (Crossref) [2023-08-02].

Spektor, M.S., Kellen, D., Hotaling, J.M., 2018. When the good looks bad: An experimental exploration of the repulsion effect. Psychological science 29, 1309–1320.

Spektor, M.S., Kellen, D., Klauer, K.C., 2022. The repulsion effect in preferential choice and its relation to perceptual choice. Cognition 225, 105164.

Trueblood, J.S., 2015. Reference point effects in riskless choice without loss aversion. Decision 2, 13–26. doi:10.1037/dec0000015.

Tversky, A., 1972. Elimination by aspects: A theory of choice.. volume 79. Psychological review.

Wedell, D.H., 1991. Distinguishing among models of contextually induced preference reversals doi:10.1037/0278-7393.17.4.767. 112 citations (Crossref) [2023-07-29].