NS01: Binary choice vs. strength-of-preference

C. E. R. Edmunds

Notes

People to email results: najmi.husaini@warwick.ac.uk

Issues

1. What does the interaction mean for rt? (normed vs. not?)

Rationale

EQ4D

average number of fixations per picture?

We are interested in the effects of attention (as measured by time spent looking at an object using eye-tracking) and value (as elicited from the subjects directly via ratings) when determining choices between two objects (such as between two bags of chips, or two pictures of nature, or between two fruits). In previous experiments, we have found that choice is predicted by an interaction of attention and value: people attend to objects they like more and are more likely to choose them. However, in other experiments there are only additively separable effects of attention and value on choice: people are more likely to pick the option they value more and/or they are more likely to choose the option they attend to more, but that these effects do not interact (i.e. looking more does not have a bigger effect when the value difference is bigger).

The current experiment will attempt to identify which properties lead to the interactive vs. additive effect of attention and value. Here, we will compare simple binary

choice between two pictures (Would you prefer Picture A or Picture B on your wall?) with a strength of preference comparison (By how much would you prefer Picture A over Picture B, or vice versa?). We hypothesise that the size of the interaction term will be greater in the strength-of-preference condition than in the choice condition.

Method

Participants

Data was collected from 67 participants: 2 of these were excluded because the eye-tracker would not initially calibrate and 12 of these were excluded due to a coding error which meant the fixation cross did not work for them. This resulted in data being collected from 53 participants. All participants were recruited from the University of Warwick's volunteer subject pool and paid £10 for their participation.

Apparatus

The participants were tested individually using an EyeLink 1000 Plus (SR Research, Osgoode, ON, Canada) eye-tracker. Monocular eye movements were recorded at 500Hz and fixations were identified by the eye tracker using velocity algorithms. The Areas of Interest were defined as a rectangle around the image position(s) on the screen. The experiment was displayed on a widescreen monitor (1920 x 1080 resolution, refresh). Participants were placed on a chin rest approximately 70cm away from the screen. Stimulus presentation was controlled by MATLAB using Psychtoolbox extensions (Brainard, 1997; Pelli, 1997).

Design

All participants completed binary choice and strength-of-preference tasks in a counterbalanced order, followed by a final valuation task where they had to rate their overall liking for each picture on a Likert scale.

Stimuli and choices

The stimuli were chosen from the International Affective Picture System (Lang, Bradley, & Cuthbert, 2008). The pictures were all positive in affect (average, male and female ratings between 5=neutral and 7=mildly positive) and had differences in value ratings of no more than 1.5 between male and female raters. After visual inspection, a further 7 images were removed for containing sexual images and 32 images were removed because they had a portrait aspect ratio. The 200 stimuli for each participant were randomly sampled without replacement from the 253 pictures that met these criteria. The participant's choices were generated by pairing the first stimulus with the hundred-and-first, the second with the hundred-and-second and so forth.

Procedure

The experiment was displayed on a black background with white text and response scales. At the beginning of the experiment the participants were asked to provide their age and gender. Then, participants completed three tasks: the binary choice task, the strength-of-preference task and the valuation task. The order of the binary choice and strength-of-preference tasks were counterbalanced between participants. For each task, the participants were shown the instructions for the task, then the eye-tracker was calibrated and then they were shown a reminder of the task instructions at which point they had to give a left mouse click to start the task. At the beginning of each trial, a fixation cross was displayed in the center of the screen until the participant had looked at it.

In both of the choice tasks, two landscape pictures (each $514x384px$) were displayed side by side after the fixation cross. The response scale was presented horizontally centered, below the stimuli. For the binary choice task, two labels ("Option A" and "Option B") were shown underneath the appropriate stimuli. The current choice was signified by a red, square marker ($30x30px$) above the label. For the strength of preference task, the response scale was a white bar displayed underneath the stimuli that extended

from the middle of one stimulus to the middle of the other. A red marker slid along the bar to signify the amount of preference for each option. The end of the scales were marked "Option A" and "Option B." In this task participants could move the marker to any point along the line using the mouse. In both tasks, the marker was initially centered equidistant between the two images. To respond in both tasks, the participants had to press the left mouse button. Reaction times were measured from the start of the trial to the beginning of the mouse click (i.e. the program did not wait for the release of the mouse button). A blank, black screen was displayed for $500ms$ between each trial.

In the final, valuation task, participants judged how much they liked each picture on a vertical Likert scale (1=strongly dislike, 7=strongly like). Each of the 200 stimuli were displayed once in a random order. Participants were offered the chance to take a self-paced break every 50 stimuli. A blank, black screen was displayed for $500ms$ between each rating. Throughout the experiment, the eye-tracker was validated every 25 trials.

Analysis

The continuous scale was split into a hundred bins. Areas of interest were defined as the area of the stimulus and a box around the response scales.

Results

Exclusions

One participant was excluded as they spent less than 60% of the time on task during the binary experiment phase.

As pre-registered, participants were excluded on a task by task basis. In previous eye-tracking research, we found that some participants spend a considerable amount of time off task, i.e. not looking at either the stimuli or the response scale. Here, only participant was found to be an outlier in the binary task and their data was removed. An outlier here is defined as the average proportion of time across all binary choice trials was

less than the first quartile of all participants minus 1.5 times the interquartile range. This left 53 participants.

We also pre-registered excluding trials for which the reaction time was less than $200ms$ or greater than the mean plus three standard deviations (this boundary was calculated across all trials). This resulted in 2.04% of trials being removed from the strength-of-preference task, and 1.23% of trials removed from the binary task. The maximum number of trials excluded for a single participant was 15. There was no difference in the models if we included the information from all trials (see Appendix).

Finally, there was one trial in the binary task where the participant was not registered as looking at either option. We removed this trial as it causes issues for calculating measures of attention.

Operationalizing value and attention

Before proceeding to the analysis, here we briefly review the measures that are used in the mixed random effects models of choice below. Value was incorporated in the models as "difference in value"

$$\Delta_V = V_{\text{right}} - V_{\text{left}} \tag{1}$$

where V_i is the value of each stimulus taken from participants' judgements in the final valuation task. Attention was incorporated in the model as

$$A = \frac{A_{\text{left}}}{A_{\text{left}} + A_{\text{right}}} \tag{2}$$

where A_i is the total time spent fixating on option i. Finally, the interaction between value and attention was defined as

$$\text{Interaction} = V_{\text{left}} A_{\text{left}} - V_{\text{right}} A_{\text{right}}. \tag{3}$$

This interaction corresponds to the interaction term in the aDDM. For a proof, please see the Appendix.

To predict reaction time, the constructs were defined similarly. For value and the interaction, we took the absolute value. For attention, we calculated

$$|A| = \left| \frac{A_{\text{left}}}{A_{\text{left}} + A_{\text{right}}} - 0.5 \right| \tag{4}$$

so that the value of $|A|$ varied between zero and a half. These recoded variables were necessary as there was no reason to predict that participants would respond faster to a particular side or particular value. Rather, we would expect participants to respond faster the greater the absolute difference between the options in both attention or value.

Predicting reaction times

There was a significant interaction between task order and condition on reaction time, $t(3) = 8.00$, $SE = 79.52$, $p < .001$. Participants who completed the continuous task first were faster than those that completed it second, whereas participants who completed the binary task first were slower than those that completed it second (see Figure 1).

There was a significant variance in reaction time intercepts across participants, $SD = 1114.18$ (95% CI: 900.45, 1378.63), $\chi^2(1) = 1375.60$, $p < .001$. Additionally, slopes based on absolute difference in attention ($|A|$) significantly varied across participants, $SD = 556.09$ (95% CI: 275.62, 1121.96), $\chi(1) = 11.35$, $p = .001$. As did the slopes based on absolute difference in value ($|\Delta_V|$), $SD = 90.54$ (95% CI: 52.81, 155.24), $\chi(1) = 8.21$, $p = .004$. Therefore, the final model included main effects of task, attention and value, with random effects of attention and value.

Table **??** contains the coefficients and 95% confidence intervals for the fixed effects in the full reaction time model for the first block (see Figure 2). In summary, reaction times were significantly predicted by absolute attention, $t = 3.05$, $SE = 477.08$, $p =$. The
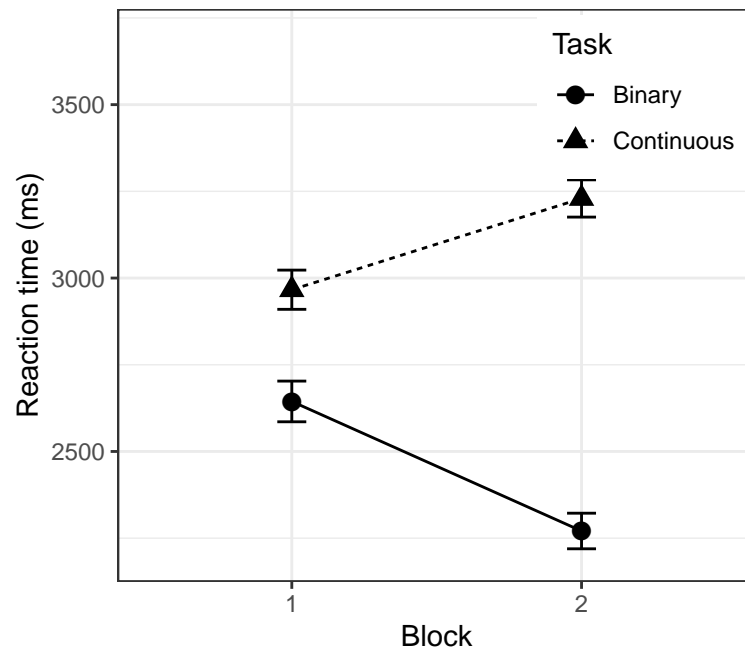
Figure 1. Mean reaction times. Error bars are within-subject difference-adjusted confidence intervals (Baguley, 2012).
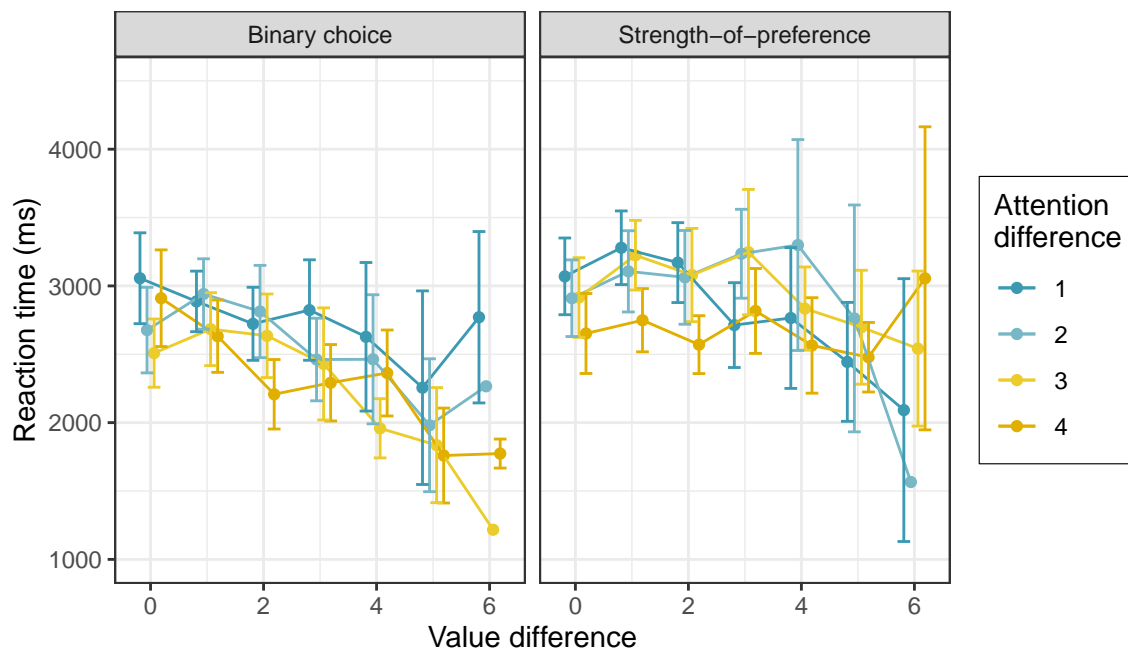

Figure 2. Looking at reaction time across difference in attention quartiles, where 1=not much difference in attention between options, 4 a lot of difference in attention between options.

Table 1. Summary of coefficients of model predicting reaction time across both blocks.

	Dependent variable:
	rt
Task	776.275*** (613.138, 939.411)
$\|A\|$	−1,000.064*** (−1,652.045, −348.084)
$\|\Delta_V\|$	−132.377*** (−190.922, −73.832)
Interaction	−5.023 (−29.935, 19.888)
Task : $\|A\|$	−1,077.605** (−1,956.334, −198.876)
Task : $\|\Delta_V\|$	10.953 (−66.664, 88.570)
Task : Interaction	3.456 (−32.076, 38.988)
TaskBinary : $\|A\|$: $\|\Delta_V\|$	160.608 (−128.490, 449.706)
TaskContinuous : $\|A\|$: $\|\Delta_V\|$	254.671* (−38.437, 547.778)
Constant	2,796.372*** (2,515.475, 3,077.269)
Observations	5,114
Log Likelihood	−43,130.270
Akaike Inf. Crit.	86,288.530
Bayesian Inf. Crit.	86,380.090

Note: $\|\Delta_A\|$ = absolute attention difference; $\|\Delta_V\|$ = absolute value difference;
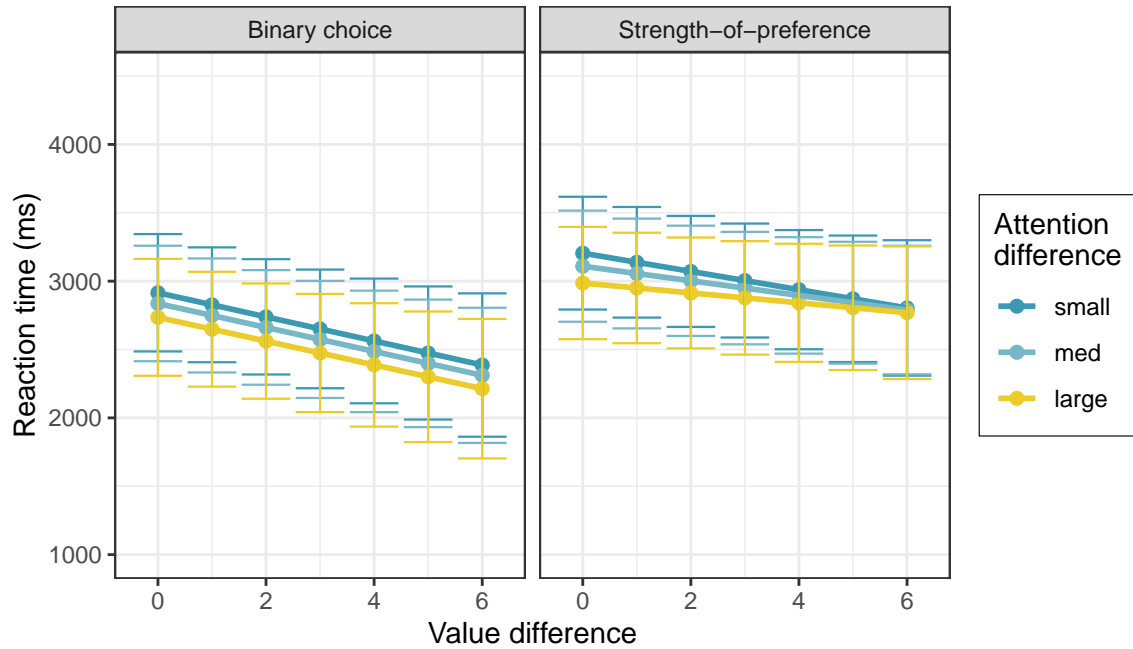

Figure 3. Predicted reaction times based on the model.

Table 2. Summary of coefficients of model predicting reaction time in Block 1.

	Dependent variable:
	rt
Task	286.288 ($-$324.899, 897.475)
$|A|$	$-$1,452.915*** ($-$2,387.976, $-$517.853)
$|\Delta_V|$	$-$89.371** ($-$175.013, $-$3.729)
Interaction	$-$21.542 ($-$59.705, 16.621)
Task : $|A|$	$-$33.801 ($-$1,307.399, 1,239.797)
Task : $|\Delta_V|$	5.426 ($-$110.689, 121.540)
Task : Interaction	27.702 ($-$23.307, 78.710)
TaskBinary : $|A|$: $|\Delta_V|$	30.806 ($-$415.297, 476.909)
TaskContinuous : $|A|$: $|\Delta_V|$	231.802 ($-$151.478, 615.081)
Constant	3,020.981*** (2,580.860, 3,461.101)
Observations	2,557
Log Likelihood	$-$21,543.170
Akaike Inf. Crit.	43,114.330
Bayesian Inf. Crit.	43,196.180

Note: $|\Delta_A|$ = absolute attention difference; $|\Delta_V|$ = absolute value difference;

greater the difference in amount of attention between the two options, the faster the participant responded. There was also an effect of absolute difference in value, $t = 2.04$, $SE = 43.70$, $p = .043$. The greater the difference in value between the choices, the faster participants responded. The main effect of task, the two-way and three-way interactions did not reach significance ($p >= 0.331$).

Choice

To consider a mixed effects model of choice, we recoded the responses given in the continuous task as binary. Other ways of completing this analysis, which attempted to take into account the additional information available in the continuous tasks (such as linear probability models), found similar results.

There was no interaction between task and order on choice, $p = .766$. Additionally, a main effects analysis found the four-way interaction was not significant, $p = .999$.

Table 3. Summary of coefficients of model predicting reaction time in Block 2.

	Dependent variable:
	rt
Task	1,139.023*** (620.872, 1,657.174)
$\lvert A \rvert$	−1,213.262*** (−2,084.689, −341.835)
$\lvert \Delta_V \rvert$	−176.758*** (−255.055, −98.462)
Interaction	−2.880 (−34.912, 29.152)
Task : $\lvert A \rvert$	−1,118.154* (−2,428.797, 192.489)
Task : $\lvert \Delta_V \rvert$	1.668 (−117.094, 120.429)
Task : Interaction	−32.883 (−81.767, 16.001)
TaskBinary : $\lvert A \rvert$: $\lvert \Delta_V \rvert$	354.851* (−13.307, 723.009)
TaskContinuous : $\lvert A \rvert$: $\lvert \Delta_V \rvert$	310.205 (−126.378, 746.788)
Constant	2,673.470*** (2,315.593, 3,031.348)
Observations	2,557
Log Likelihood	−21,457.680
Akaike Inf. Crit.	42,943.360
Bayesian Inf. Crit.	43,025.220

Note: $\lvert \Delta_A \rvert$ = absolute attention difference; $\lvert \Delta_V \rvert$ = absolute value difference;

However, there was a significant effect of task order on reaction time.

There was a significant random effect of difference in value on choice, $\chi^2(3) = 2081.55$, $p < .001$. Adding random slopes of attention significantly added to the model, $\chi^2(1) = 857.88$, $p < .001$, whereas also adding a random effect of task did not, $\chi^2(3) = 0.67$, $p = .881$. However, adding random slopes based on the attention resulted in the models not converging, so we drop this term in the following models.

Table 8 contains the coefficients and 95% confidence intervals for the fixed effects in the full choice model for the first block only (the preregistered analysis). Choice was significantly predicted by attention, $t(1) = 12.89$, $SE = 0.44$, $p < .001$, and value, $t(1) = 14.28$, $SE = 0.06$, $p < .001$. In other words, the picture that was looked at more was more likely to be picked, as was the picture that the participant preferred. The main effect of task, as well as two- and three-way interactions did not reach significance ($p >= .168$). The predicted choice probabilities are shown in Figure 6.
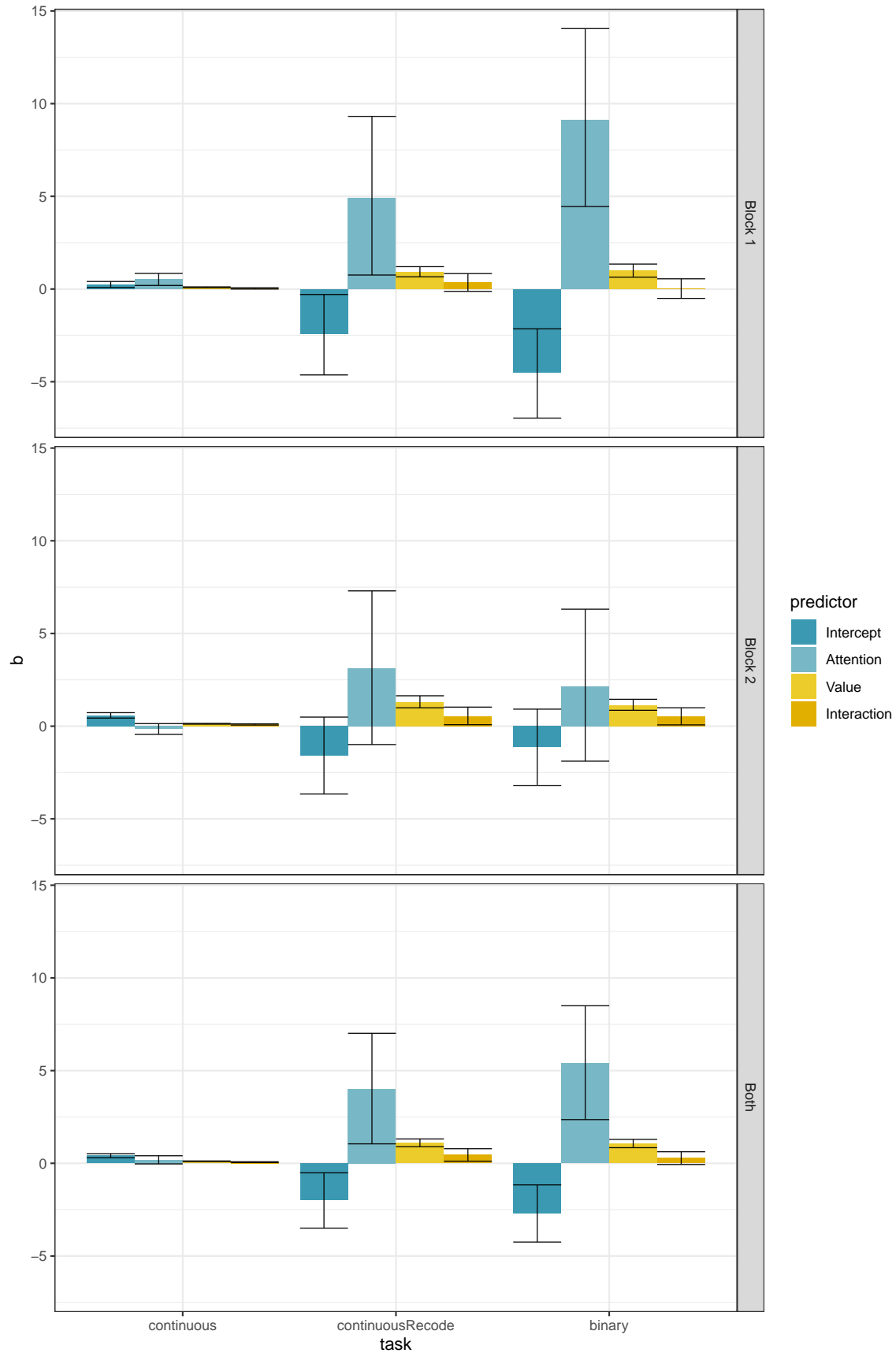
Figure 4. Parameter estimates with unstandardised coefficients. Error bars are 95% confidence intervals.
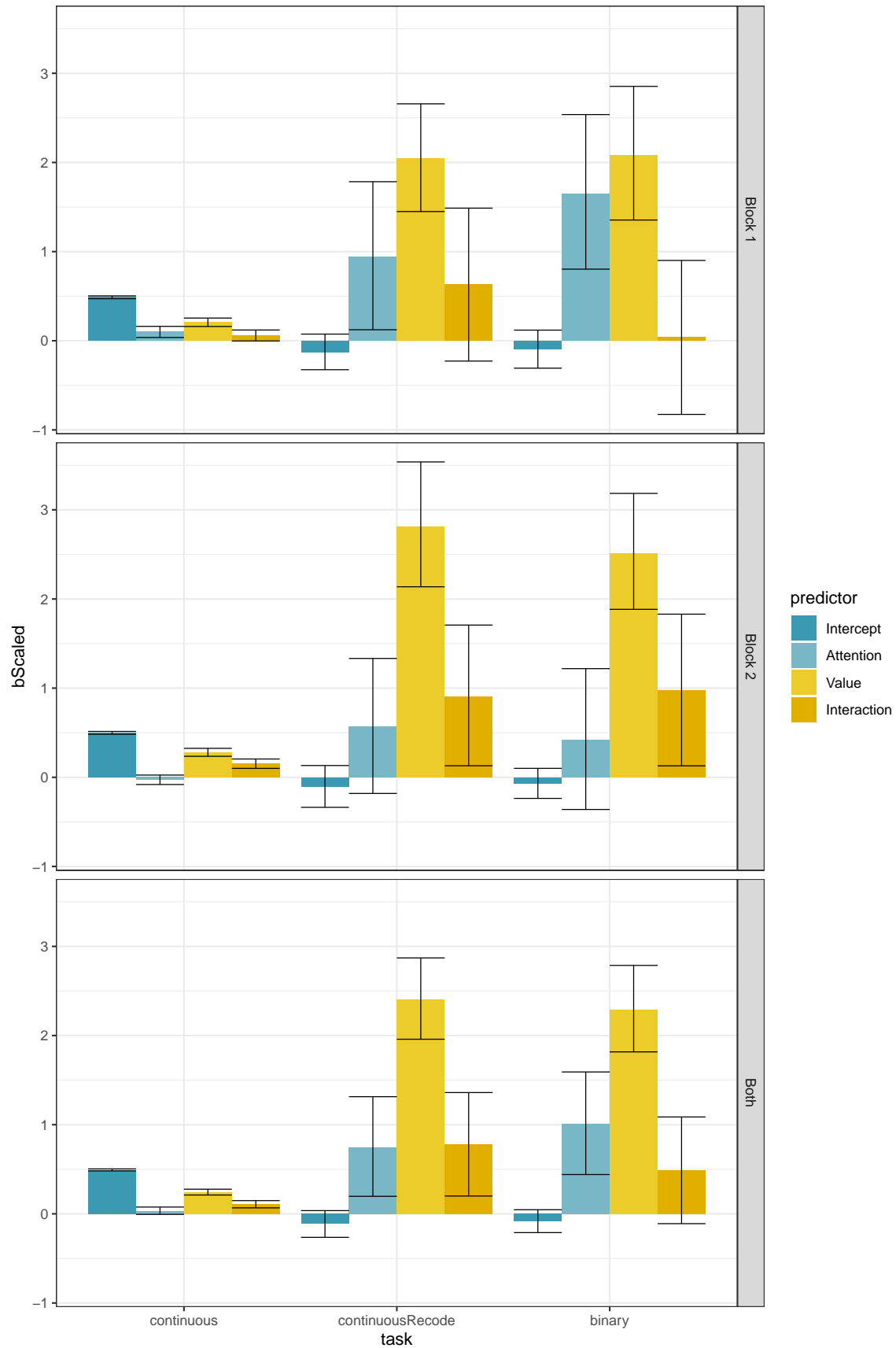
Figure 5. Parameter estimates with standardised coefficients. Error bars are 95% confidence intervals.

Also included are the outputs for both blocks together (Table 7) and for the final block on its own (Table 9). The final block version is perhaps interesting as the reaction time analyses showed that the difference in reaction was greater for the second block than the first. Although this did change the coefficients slightly, the confidence intervals are still

Table 4. Summary of coefficients of model predicting choice (both blocks).

	Dependent variable:	
	Choice	
	Strength-of-prefence	Binary choice
	(1)	(2)
Attention	3.999*** (1.028, 6.969)	5.382*** (2.328, 8.436)
Δ_V	1.101*** (0.894, 1.308)	1.060*** (0.838, 1.282)
Interaction	0.450*** (0.116, 0.785)	0.279 (−0.063, 0.620)
Constant	−1.987*** (−3.473, −0.501)	−2.683*** (−4.215, −1.151)
Observations	2,547	2,567
Log Likelihood	−963.245	−973.267
Akaike Inf. Crit.	1,938.490	1,958.535
Bayesian Inf. Crit.	1,973.546	1,993.638

Note: Δ_V = value difference;

Table 5. Summary of coefficients of model predicting choice by task (Block 1).

	Dependent variable:	
	Choice	
	Strength-of-prefence	Binary choice
	(1)	(2)
Attention	4.890** (0.570, 9.211)	9.121*** (4.345, 13.897)
Δ_V	0.926*** (0.654, 1.198)	0.980*** (0.632, 1.327)
Interaction	0.353 (−0.125, 0.832)	0.023 (−0.503, 0.549)
Constant	−2.421** (−4.583, −0.259)	−4.491*** (−6.887, −2.094)
Observations	1,326	1,231
Log Likelihood	−517.458	−436.827
Akaike Inf. Crit.	1,046.916	885.655
Bayesian Inf. Crit.	1,078.056	916.348

Note: Δ_V = value difference;

securely including 0.

Interaction affected by task

Table 6. Summary of coefficients of model predicting choice by task (Block 2).

	Dependent variable:	
	Choice	
	Strength-of-prefence	Binary choice
	(1)	(2)
Attention	3.125 (−0.994, 7.245)	2.154 (−1.924, 6.233)
Δ_V	1.299*** (0.981, 1.618)	1.144*** (0.852, 1.436)
Interaction	0.547** (0.074, 1.020)	0.528** (0.068, 0.988)
Constant	−1.572 (−3.634, 0.490)	−1.109 (−3.154, 0.937)
Observations	1,221	1,336
Log Likelihood	−442.390	−530.581
Akaike Inf. Crit.	896.779	1,073.162
Bayesian Inf. Crit.	927.424	1,104.347

Note: Δ_V = value difference;

Table 7. Summary of coefficients of model predicting choice (both blocks).

	Dependent variable:
	recodedResponse
Task	−0.013 (−0.178, 0.152)
Δ_A	5.677*** (4.814, 6.540)
Δ_V	0.891*** (0.769, 1.014)
Task : Δ_A	0.574 (−0.242, 1.389)
Task : Δ_V	0.022 (−0.084, 0.128)
Δ_A : Δ_V	0.060 (−0.281, 0.401)
Task : Δ_A : Δ_V	−0.112 (−0.618, 0.395)
Constant	0.017 (−0.115, 0.150)
Observations	5,115
Log Likelihood	−1,881.974
Akaike Inf. Crit.	3,785.949
Bayesian Inf. Crit.	3,857.888

Note: Δ_A = attention difference; Δ_V = value difference;

Table 8. Summary of coefficients of model predicting choice (Block 1 only).

	Dependent variable:
	recodedResponse
Task	−0.073 (−0.383, 0.237)
Δ_A	6.556*** (5.137, 7.976)
Δ_V	0.933*** (0.746, 1.119)
Task : Δ_A	0.127 (−1.812, 2.067)
Task : Δ_V	−0.129 (−0.373, 0.114)
Δ_A : Δ_V	0.025 (−0.555, 0.604)
Task : Δ_A : Δ_V	−0.025 (−0.795, 0.746)
Constant	0.068 (−0.157, 0.293)
Observations	2,557
Log Likelihood	−931.833
Akaike Inf. Crit.	1,885.665
Bayesian Inf. Crit.	1,949.978

Note: Δ_A = attention difference; Δ_V = value difference;

Table 9. Summary of coefficients of model predicting choice (Block 2 only).

	Dependent variable:
	recodedResponse
Task	−0.004 (−0.259, 0.252)
Δ_A	5.199*** (3.984, 6.414)
Δ_V	0.909*** (0.727, 1.090)
Task : Δ_A	0.974 (−0.809, 2.758)
Task : Δ_V	0.106 (−0.154, 0.365)
Δ_A : Δ_V	0.071 (−0.368, 0.509)
Task : Δ_A : Δ_V	−0.195 (−0.951, 0.561)
Constant	−0.017 (−0.190, 0.156)
Observations	2,558
Log Likelihood	−961.365
Akaike Inf. Crit.	1,944.730
Bayesian Inf. Crit.	2,009.047

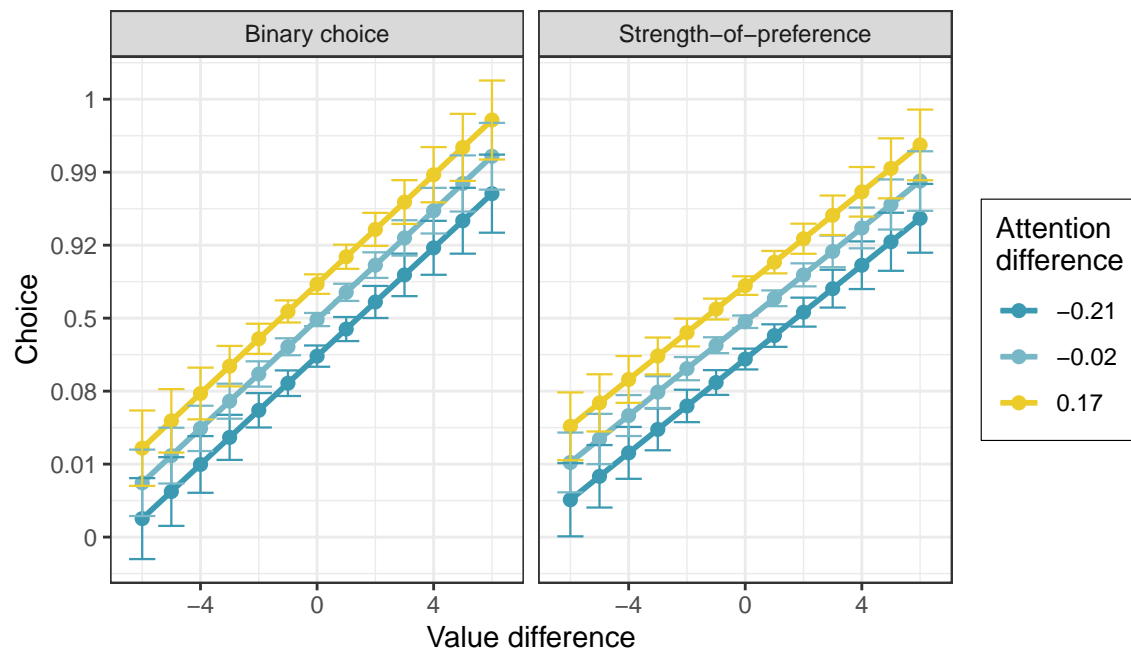*Note:* Δ_A = attention difference; Δ_V = value difference;

Figure 6. Predicted choice probabilities from the model, based on the first block only.

Overall value

Smith and Krajbich (2018) looked at overall value along with value difference to predict reaction time using the model:

$$\log(RT) \sim \beta_0 + \beta_1 |V_{\text{left}} - V_{\text{right}}| + \beta_2 (V_{\text{left}} + V_{\text{right}}) \tag{5}$$

where V_i is the subjective value of option i. I do a similar thing below. Using this model, we fail to replicate their findings. In this experiment, there is a main effect of absolute difference in value (Δ_V), $\beta_1 = -75.24$, $t = 3.53$, $SE = 21.32$, $p < .001$. In other words, the greater the difference in value between the choices, the faster participants were. However, there is no main effect of total value, $\beta_2 = 11.39$, $t = 0.85$, $SE = 13.36$, $p = .394$.

Of course, using this model is silly because this collapses across tasks. Therefore, we also ran a linear model that looked at Δ_V, total value and task on reaction time in the first

Table 10. Summary of coefficients of model predicting reaction time from total value and value difference

	Dependent variable:
	rt
Task	720.525** (14.210, 1,426.840)
$\sum_A$	41.222 (−13.101, 95.545)
Δ_V	−90.698 (−379.134, 197.737)
Task : $\sum_A$	−60.676 (−138.112, 16.760)
Task : Δ_V	60.638 (−341.784, 463.060)
$\sum_A : \Delta_V$	−3.902 (−37.350, 29.547)
Task : $\sum_A : \Delta_V$	2.386 (−44.019, 48.790)
Constant	2,490.447*** (2,001.486, 2,979.407)
Observations	2,557
R^2	0.021
Adjusted R^2	0.018
Residual Std. Error	1,479.557 (df = 2549)
F Statistic	7.728*** (df = 7; 2549)
Note:	$\sum_A$ = total value; Δ_V = value difference;

block (see Table 10). Here, we just found a main effect of task, $t = 2.00$, $SE = 360.37$, $p = .046$.

Open Practices Statement

The analysis was pre-registered at `AsPredicted.org/19698`. The data and analysis of this experiment is available in the Open Science Framework `https://osf.io/xfc8a/`.

Appendix

Derivation of interaction term

Choice model including all trials

Here, we consider the model of choice including all trials (i.e. no exclusions based on reaction time) for the first block. The model for the second block is similar.

There was a significant random effect of difference in attention on choice, $\chi^2(1) = 985.41$, $p < .001$. There was also a random effect of difference in value on choice, $\chi^2(1) = 569.79$, $p < .001$. A random effect of task did not add additional information, $\chi^2(3) = 0.02$, $p = .999$. Therefore, the final model included fixed effects of task, attention and value, with random effects of attention and value.

Table **??** contains the coefficients and 95% confidence intervals for the fixed effects in the full choice model. Choice was significantly predicted by attention, $t(1) = 8.90$, $SE = 0.72$, $p < .001$, and value, $t(1) = 9.73$, $SE = 0.10$, $p < .001$. In other words, the

Table 11. Summary of coefficients of model predicting choice including all trials

	Dependent variable:
	recodedResponse
Task	-0.083 (-0.382, 0.216)
Δ_A	6.436^{***} (5.019, 7.853)
Δ_V	0.926^{***} (0.740, 1.113)
Task : Δ_A	0.446 (-1.486, 2.379)
Task : Δ_V	-0.146 (-0.388, 0.096)
Δ_A : Δ_V	0.059 (-0.513, 0.631)
Task : Δ_A : Δ_V	-0.039 (-0.793, 0.714)
Constant	0.061 (-0.157, 0.280)
Observations	2,650
Log Likelihood	-969.822
Akaike Inf. Crit.	1,961.644
Bayesian Inf. Crit.	2,026.349
Note:	Δ_A = attention difference; Δ_V = value difference;

picture that was looked at more was more likely to be picked, as was the picture that the participant preferred. The main effect of task, as well as two- and three-way interactions did not reach significance ($p >= .237$).

References

Baguley, T. (2012). Calculating and graphing within-subject confidence intervals for ANOVA. *Behavior Research Methods*, *44*, 158–175.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* (Tech. Rep.). Gainesville, FL: University of Florida.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Smith, S. M., & Krajbich, I. (2018, December). Gaze Amplifies Value in Decision Making. *Psychological Science*, *30*(1), 116–128.