

# NS01: Binary choice vs. strength-of-preference

C. E. R. Edmunds

## Notes

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

## Rationale

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 Warwicks 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  $514 \times 384px$ ) 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 ( $30 \times 30px$ ) 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.

### Operationalizing attention and value

Before proceeding to the analysis, a quick note on the measures we constructed and used in the mixed random effects models below. Value was incorporated in the models as “difference in value”

$$\Delta_V = V_{\text{right}} - V_{\text{left}} \quad (1)$$

where  $V_i$  is the value of each stimulus taken from participants’ judgements in the final valuation task. Attention was incorporated in the models as “difference in attention”

$$\Delta_A = \frac{A_{\text{right}} - A_{\text{left}}}{RT} \quad (2)$$

where  $A_i$  is the total time spent fixating on option  $i$  and  $RT$  is the reaction time.<sup>1</sup>

In the models of reaction time,  $\Delta_A$  and  $\Delta_V$  were similarly defined, although the absolute value of each construct was taken as there is no reason to predict that participants would respond faster to the left (or right). Rather, we would expect participants to respond faster the greater the absolute difference between the options.

### Predicting reaction times

There was a significant interaction between task order and condition on reaction time,  $t(3) = 8.01$ ,  $SE = 79.51$ ,  $p < .001$ . 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).

Further, the four-way interaction of task, attention, value with task order was marginal,

---

<sup>1</sup>Note that other ways of normalizing attention such as time spent on task or time spent looking at the two options made no difference. See section below.

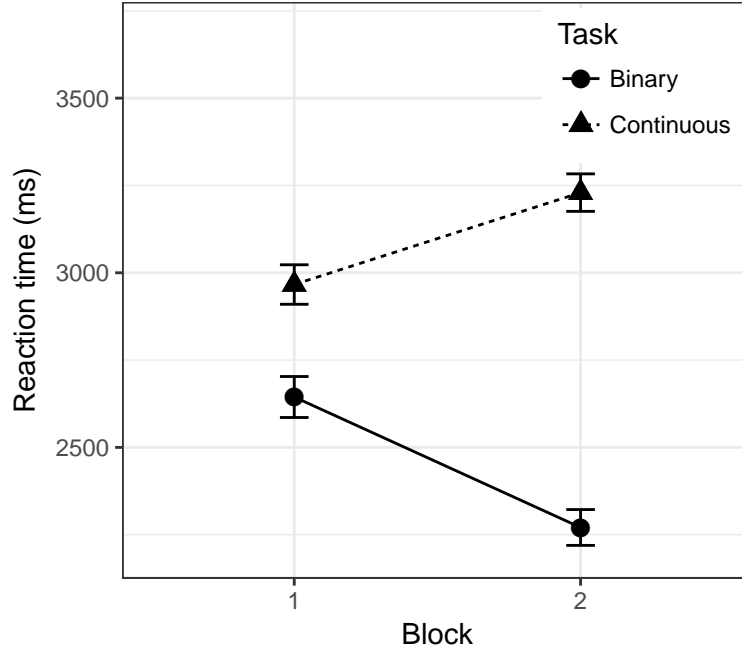


Figure 1. Mean reaction times. Error bars are within-subject difference-adjusted confidence intervals (Baguley, 2012).

$t = 1.88$ ,  $SE = 139.16$ ,  $p = .060$ . So, as preregistered, in the following we only considered information from the first block.

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 ( $|\Delta_A|$ ) significantly varied across participants,  $SD = 556.09$  (95% CI: 275.62, 1121.96),  $\chi(1) = 8.21$ ,  $p = .004$ . As did the slopes based on absolute difference in value ( $|\Delta_V|$ ),  $SD = 90.54$  (95% CI: 52.81, 155.24),  $\chi(1) = 8.34$ ,  $p = .004$ . Therefore, the final model included main effects of task, attention and value, with random effects of attention and value.

Table 1 contains the coefficients and 95% confidence intervals for the fixed effects in the full reaction time model (see Figure 2). In summary, reaction times were significantly predicted by difference in attention,  $t = 2.44$ ,  $SE = 318.49$ ,  $p = .015$ . The greater the difference in amount of attention between the two options, the faster the participant responded. There was also an effect of difference in value,  $t = 2.03$ ,  $SE = 43.70$ ,  $p = .043$ .

Table 1. Summary of coefficients of model predicting reaction time

	Dependent variable:
	rt
Task	304.578 (−308.454, 917.611)
$ \Delta_A $	−778.436** (−1,402.672, −154.200)
$ \Delta_V $	−88.528** (−174.174, −2.882)
Task : $ \Delta_A $	−168.390 (−1,017.792, 681.012)
Task : $ \Delta_V $	10.001 (−105.245, 125.247)
$ \Delta_A $ : $ \Delta_V $	4.835 (−291.029, 300.699)
Task : $ \Delta_A $ : $ \Delta_V $	127.742 (−268.569, 524.053)
Constant	2,984.306*** (2,541.827, 3,426.784)
Observations	2,557
Log Likelihood	−21,556.860
Akaike Inf. Crit.	43,137.720
Bayesian Inf. Crit.	43,207.880

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

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

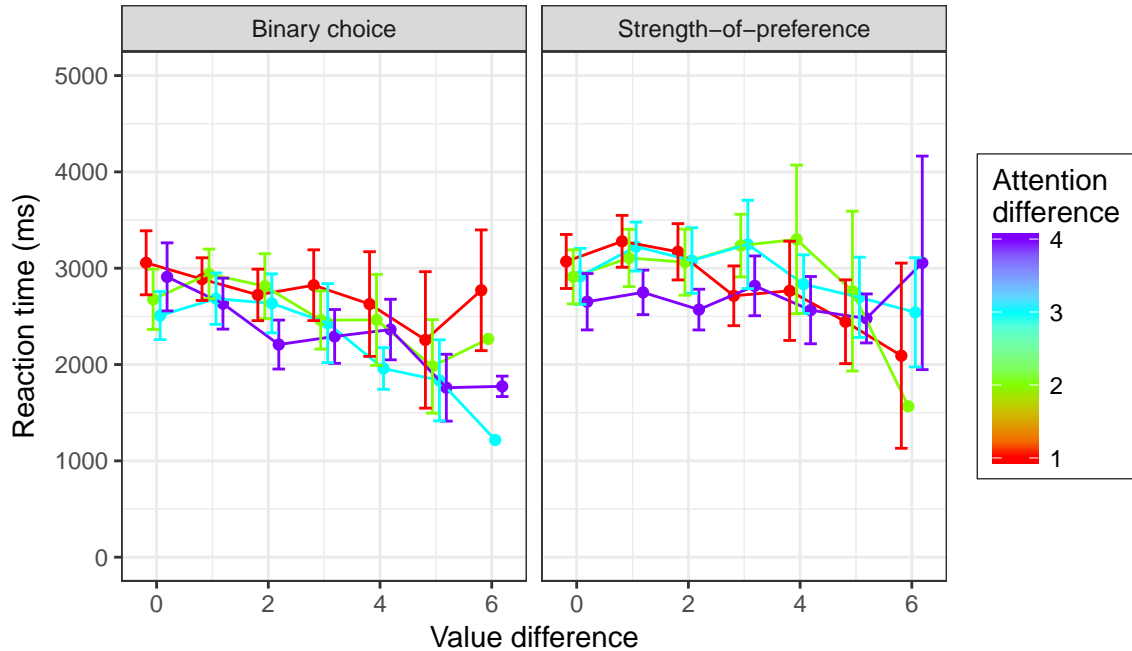


Figure 2. Looking at reaction time across difference in attention quartiles, where 1=not much difference in attention between options, 4 a lot of difference in attention between options.

( $p \geq 0.331$ ).

## 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,  $p = .999$ . However, there was a significant effect of task order on reaction time, so here we report an analysis of the first block only.

There was a significant random effect of difference in attention on choice,  $\chi^2(1) = 1791.02$ ,  $p < .001$ . There was also a random effect of difference in value on choice,  $\chi^2(1) = 1341.70$ ,  $p < .001$ . A random effect of task did not add additional information,  $\chi^2(3) = 0.46$ ,  $p = .928$ . Therefore, the final model included fixed effects of task, attention

Table 2. Summary of coefficients of model predicting choice

	<i>Dependent variable:</i>
	recodedResponse
Task	0.014 (−0.495, 0.523)
$\Delta_A$	5.928*** (3.520, 8.336)
$\Delta_V$	1.007*** (0.749, 1.266)
Task : $\Delta_A$	0.416 (−2.710, 3.541)
Task : $\Delta_V$	−0.268* (−0.583, 0.047)
$\Delta_A$ : $\Delta_V$	−0.270 (−1.320, 0.781)
Task : $\Delta_A$ : $\Delta_V$	0.571 (−0.693, 1.834)
Constant	−0.129 (−0.523, 0.265)
Observations	1,009
Log Likelihood	−376.774
Akaike Inf. Crit.	775.547
Bayesian Inf. Crit.	829.631

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



and value, with random effects of attention and value.

Table 2 contains the coefficients and 95% confidence intervals for the fixed effects in the full choice model. 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 ( $p \geq .168$ ).

### Alternative analyses

So, the above (pre-registered) analyses indicate no substantial differences between the binary choice and strength-of-preference tasks. Some remaining possibilities are that the *pattern* of attention is different between these two tasks. Possible metrics that might capture these differences are:

#### Number of fixations

Here, we operationalized “number of fixations” as

$$\frac{N_{right}}{N_{left} + N_{right}} \quad (3)$$

where  $N_i$  is the number of fixation on option  $i$ .

There was significant variance of participant on proportion of number of fixations,  $\chi^2(1) = 25.97$ ,  $p < .001$ . Therefore, the model included fixed effects of task, attention and value with their interactions and a random effect of participant on the intercepts.

As shown in Table 3, nothing really came out.

#### Duration of fixations

Here, duration of fixations was operationalized as

$$\frac{D_{right}}{D_{left} + D_{right}} \quad (4)$$

where  $D_i$  was the mean duration of fixations on option  $i$ .

Including random effects did not improve the explanatory power of this model. Further, as seen in Table 4, nothing came out.

### Looking at different ways of operationalizing attention

In the above analyses, attention was operationalized as

*Table 3.* Summary of coefficients of model predicting number of fixations on each choice

	<i>Dependent variable:</i>
	relativeFixN
Task	−0.013 (−0.042, 0.015)
$\Delta_A$	−0.012 (−0.079, 0.054)
$\Delta_V$	0.007 (−0.002, 0.015)
Task : $\Delta_A$	0.013 (−0.073, 0.098)
Task : $\Delta_V$	−0.003 (−0.013, 0.008)
$\Delta_A$ : $\Delta_V$	0.00005 (−0.031, 0.031)
Task : $\Delta_A$ : $\Delta_V$	0.022 (−0.018, 0.061)
Constant	0.467*** (0.445, 0.489)
Observations	1,009
Log Likelihood	311.125
Akaike Inf. Crit.	−602.250
Bayesian Inf. Crit.	−553.083

*Note:*  $\Delta_A$  = attention difference;  $\Delta_V$  = value difference;

*Table 4.* Summary of coefficients of model predicting duration of time on options

	<i>Dependent variable:</i>
	relativeFixDur
Task	0.003 (−0.013, 0.019)
$\Delta_A$	0.0004 (−0.046, 0.047)
$\Delta_V$	−0.003 (−0.009, 0.003)
Task : $\Delta_A$	0.024 (−0.035, 0.084)
Task : $\Delta_V$	0.002 (−0.006, 0.009)
$\Delta_A$ : $\Delta_V$	0.008 (−0.014, 0.029)
Task : $\Delta_A$ : $\Delta_V$	−0.001 (−0.029, 0.027)
Constant	0.527*** (0.514, 0.539)
Observations	992
R <sup>2</sup>	0.004
Adjusted R <sup>2</sup>	−0.003
Residual Std. Error	0.121 (df = 984)
F Statistic	0.604 (df = 7; 984)

*Note:*  $\Delta_A$  = attention difference;  $\Delta_V$  = value difference;

$$\frac{T_{right} - T_{left}}{RT} \quad (5)$$

where  $T_i$  is the total time spent fixating on option  $i$  and  $RT$  is the reaction time.

However, previously Tim has used

$$\frac{T_{right} - T_{left}}{T_{right} + T_{left}} \quad (6)$$

where attention is normalized out of the total time spent looking at options.

Also, CERE noticed that participants appear to spend more time looking at the Likert scale in the strength-of-preference task than in the binary task. Especially when the participants appeared to be undecided between the two options. Therefore,

$$\frac{T_{right} - T_{left}}{T_{right} + T_{left} + T_{likert}} \quad (7)$$

is a measure of difference of attention normalised by time spent on task.

Mainly because (as shown in Figure 3) they all correlate pretty perfectly.

## Discussion

### Open Practices Statement

The analysis was pre-registered at [AsPredicted.org/19698](https://aspredicted.org/19698). The data and analysis of this experiment is available in the Open Science Framework <https://osf.io/xfc8a/>.

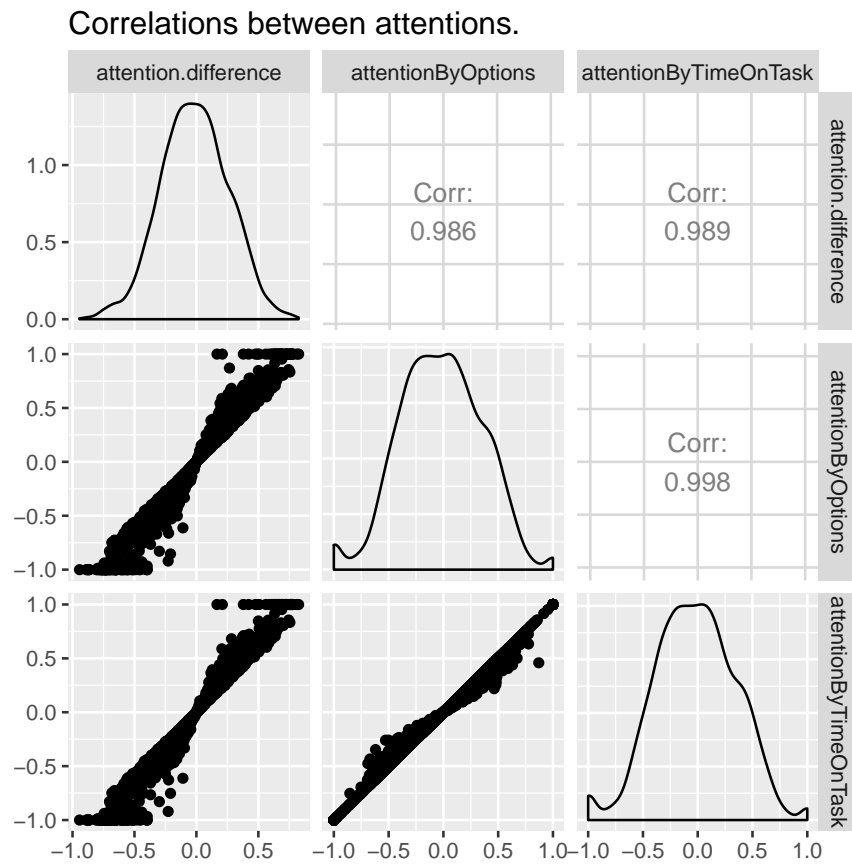


Figure 3. Correlations between different metrics of attention

## 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.

Table 5. Summary of coefficients of model predicting choice, comparing different attentions

	<i>Dependent variable:</i>		
	by reaction time (1)	recodedResponse by time on choices (2)	by time on task (3)
Task	−0.013 (−0.178, 0.152)	−0.016 (−0.182, 0.151)	−0.014 (−0.179, 0.151)
Attention	5.677*** (4.814, 6.540)	4.210*** (3.574, 4.846)	4.273*** (3.624, 4.922)
Value	0.891*** (0.769, 1.014)	0.887*** (0.763, 1.011)	0.888*** (0.765, 1.010)
Task:Attention	0.574 (−0.242, 1.389)	−0.004 (−0.588, 0.581)	0.250 (−0.345, 0.845)
Task:Value	0.022 (−0.084, 0.128)	0.029 (−0.079, 0.137)	0.027 (−0.079, 0.133)
Attention:Value	0.060 (−0.281, 0.401)	0.060 (−0.201, 0.321)	0.055 (−0.200, 0.311)
Task:Attention:Value	−0.112 (−0.618, 0.395)	−0.134 (−0.504, 0.235)	−0.127 (−0.498, 0.243)
Constant	0.017 (−0.115, 0.150)	0.014 (−0.119, 0.146)	0.015 (−0.116, 0.146)
Observations	5,115	5,114	5,114
Log Likelihood	−1,881.974	−1,895.954	−1,885.168
Akaike Inf. Crit.	3,785.949	3,813.908	3,792.336
Bayesian Inf. Crit.	3,857.888	3,885.845	3,864.273

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01