Supplementary materials

Supplementary Notes

1. Calibration & validation

   For subjects who passed the initial calibration, 67% (33 out of 49) of them passed the calibration at their first attempt, 23% (11 out of 49) of them passed the calibration at their second attempt, and 10% (5 out of 49) need the third attempt. However, there's no evidence that the initial calibration performance affected the later spatial accuracy (quantified by the averaged hit ratio across the intertrial validation dots; $F(2, 45) = 0.62$, $p = 0.54$).

2. Relationship between rating difference vs. RT and choice

   We examined the relationship between value differences and choice probability, and response times (See Fig. S1). Consistent with Krajbich et al. (2010), we found that response time and choice accuracy are functions of the choice difficulty (mixed effects regression of choice probability on value differences: $\beta = 0.63$, $p = 10^{-16}$ for the Mturk study). However, response times in the MTurk study were significantly shorter than Krajbich et al (2010) as we detailed in the main text.

Supplementary Tables

Calibration + validation attempts	Pixel level (.px)	Threshold	Subject proportion
1	130	80%	0.67
2	165	70%	0.23
3	200	60%	0.10

Supplementary Table 1. This table summarized the statistics related to three different initial calibration + validation attempts. Pixel level represents the maximum Euclidean distance between the prediction and gazed validation dot. Threshold represent the smallest proportion of the valid dots out of all validation dots in order to pass the initial calibration. Subject proportion represents proportion of the subjects who pass the calibration at the corresponding attempt.

Hit ratio	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
Mean	0.87	0.73	0.74	0.74	0.67	0.87	0.79	0.74	0.75	0.70
median	0.97	0.80	0.84	0.79	0.74	0.96	0.86	0.86	0.84	0.78

Supplementary Table 2. This table summarizes the statistics related to intertrial validation dots. There were 10 intertrial validations in total. For each validation, there were three dots, and for each dot there were multiple gaze measurements (approximately 150). For each subject, we computed the proportion of hits for each validation. Mean hit ratio represents the average of that measure, across subjects. Median hit ratio represents the median of that measure across subjects.

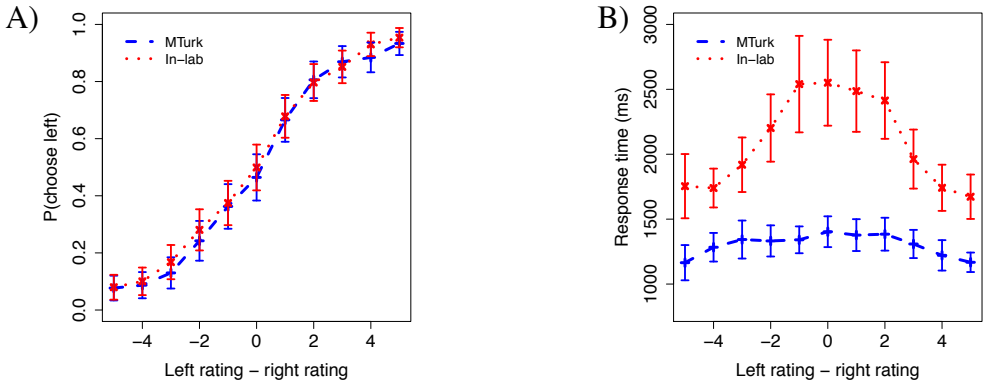
	Pass rate (all subjects)	Pass rate (> 0.45)	Pass rate (0.3 – 0.45)	Pass rate (< 0.3)
Mean (SD)	0.6 (0.26)	0.73 (0.16)	0.36 (0.04)	0.15 (0.08)
Number of subjects	49	35	8	6

Supplementary Table 3. This table summarizes the statistics related to intertrial validation pass rates. The first column shows the mean (standard deviation) pass rate for all subjects. The second to fourth columns summarize the mean (standard deviation) pass rates for subjects who met each cutoff.

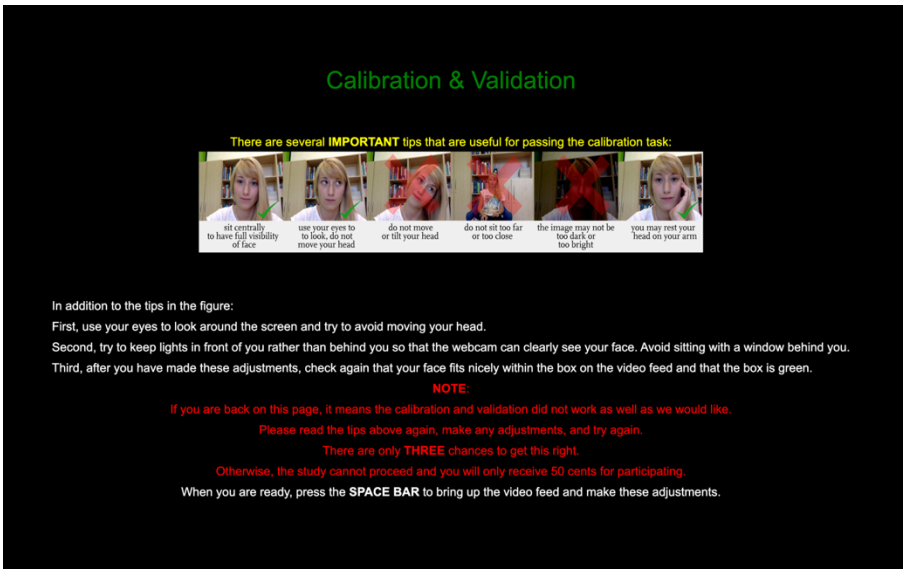
	Online	Krajbich et al. (2010)
First	344 (234)	432 (376)
Middle	312 (220)	730 (579)
Last	341 (216)	544 (543)

Supplementary Table 4. This table summarizes the statistics related to dwell times. The second and the third column display the mean (standard deviation) for first dwell times, middle dwell times, and last dwell times in the Online MTurk data and Krajbich et al. (2010) data.

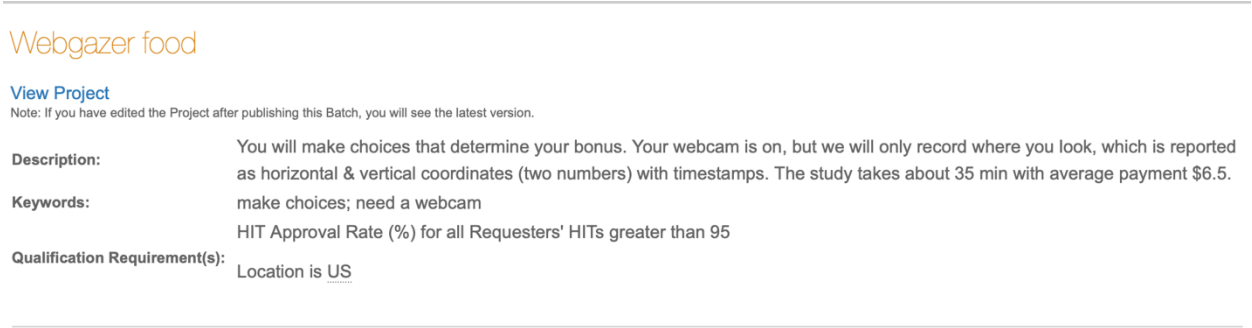
Supplementary Figures



Supplementary Figure 1. A). relationship between choice accuracy and value difference. B). relationship between response times and value difference. In each plot, the red line/dots represent the results in Krajbich et al. (2010)'s dataset; the blue line/dots represent the results in the current online MTurk study.



Supplementary Figure 2. Eye-tracking instructions shown to the subjects.

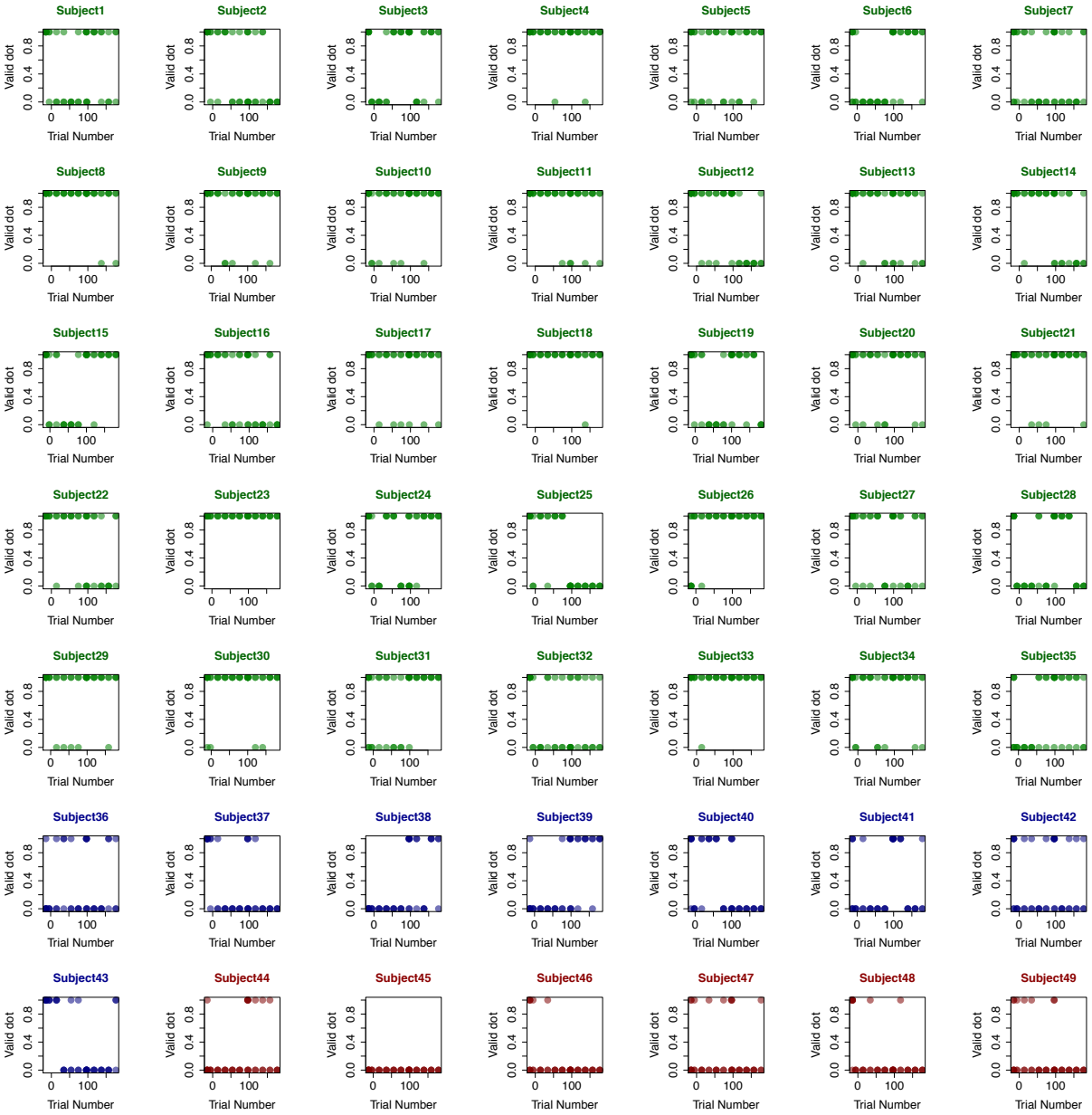


Supplementary Figure 3. MTurk experiment description and requirements.

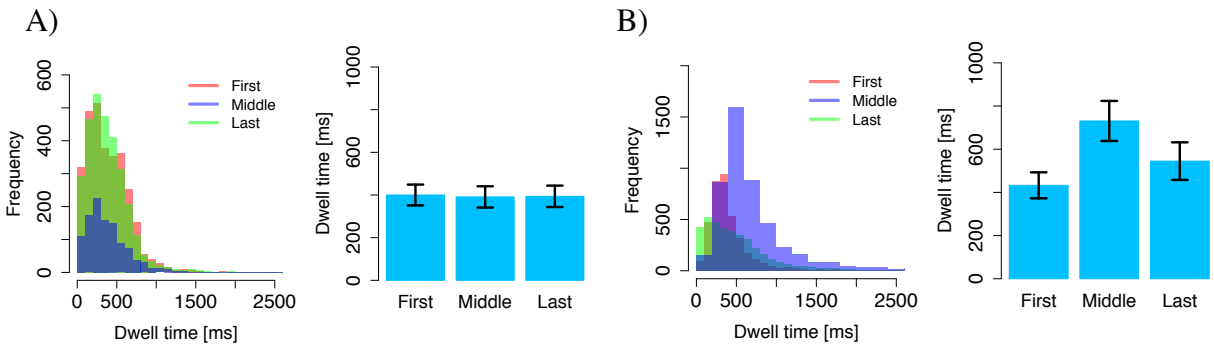
```
var inital_eye_calibration = {
  timeline: [
    eyeTrackingNote, // instruction
    {
      type: "eye-tracking",
      doInit: true,
      doCalibration: true,
      doValidation: true,
      calibrationDots:  13,
      calibrationDuration: 3,
      doValidation: true,
      validationDots:  13,
      validationDuration: 2,
      validationTol: 130,
     ],
};
```

Supplementary Figure 4. An instance of initializing the eye-tracking process with the template.

Supplementary Figure 5. Visualization of the calibration + validation task.



Supplementary Figure 6. Intertrial validation dot success (1 means the dot is valid, and 0 means the dot is invalid) as a function of trial number at the individual level. Subjects colored green had intertrial validation rates above 0.45; subjects colored blue had intertrial validation rates above 0.2; subjects colored red had intertrial validation rates below 0.2; The data from subjects colored green (1-35) were included in the analysis. Partial data from subjects 39, 40, and 43 were also included in the analysis (see methods for more details). Data from the rest of the subjects were excluded from the analysis.



Supplementary Figure 7. Dwell time distributions for (A) the online MTurk data, and (B) the Krajbich et al. 2010 data. The histograms display the distributions of first dwell times, middle dwell times, and last dwell times. The bar plots display the mean dwell time for each of those three categories.