Published in final edited form as:
Vis cogn. 2012 January 1; 20(9): 1110–1129. doi:10.1080/13506285.2012.735719.

Heuristics and Criterion Setting during Selective Encoding in Visual Decision-Making: Evidence from Eye Movements

Elizabeth R. Schotter, **Cainen Gerety**, and **Keith Rayner**
University of California, San Diego

Abstract

When making a decision, people spend longer looking at the option they ultimately choose compared other options—termed the *gaze bias effect*—even during their first encounter with the options (Glaholt & Reingold, 2009a, 2009b; Schotter, Berry, McKenzie & Rayner, 2010). Schotter et al. (2010) suggested that this is because people selectively encode decision-relevant information about the options, on-line during the first encounter with them. To extend their findings and test this claim, we recorded subjects' eye movements as they made judgments about pairs of images (i.e., which one was taken more recently or which one was taken longer ago). We manipulated whether both images were presented in the same color content (e.g., both in color or both in black-and-white) or whether they differed in color content and the extent to which color content was a reliable cue to relative recentness of the images. We found that the magnitude of the gaze bias effect decreased when the color content cue was not reliable during the first encounter with the images, but no modulation of the gaze bias effect in remaining time on the trial. These data suggest people do selectively encode decision-relevant information on-line.

Keywords

Eye Movements; Decision Making; Gaze Bias; Selective Encoding; Heuristics

It seems intuitive that people might, when faced with a decision, spend more time contemplating the option they ultimately choose compared to another option. Furthermore, the most obvious intuition suggests that this tendency yields an increasing probability of returning to and continuing to consider an option that is still "in the running" to be chosen. In fact, some theories of decision-making, such as Decision Field Theory (Busmeyer & Townsend, 1993) and Sequential Value Matching (Johnson & Busemeyer, 2005) posit that options are sampled and returned to, often many times and, later, evaluated, often with paired comparisons before a decision is ultimately made. However, recent research has found that this bias toward the ultimately chosen item can actually arise quite early in the decision-making process (Glaholt & Reingold, 2009a, 2009b; Schotter, Berry, McKenzie & Rayner, 2010) and may reflect the fact that people engage actively in *selective encoding*—actively evaluating options (e.g., putting more or less emphasis on certain option properties) during encoding and excluding options that don't fit the given criteria (Glaholt & Reingold, 2009a, 2009b, 2011; Schotter et al., 2010).

Recently, eyetracking methods have been used to study the cognitive processes underlying decision-making (see Glaholt & Reingold, 2011 for a review). The advantage of using eyetracking is that it provides information about where the subject is looking with

Correspondence to: Elizabeth R. Schotter, Department of Psychology, University of California, San Diego, La Jolla, CA 92093-0109, Phone: (858) 822-7813, Fax: (858) 534-7190, eschotter@ucsd.edu.

millisecond-to-millisecond accuracy (Rayner, 1998, 2009), thereby providing a wealth of data and the ability to analyze the decision-making process at various time points and substages. From this work, researchers have confirmed the aforementioned intuition that, when making decisions we exhibit a *gaze bias*—we often tend to spend more time looking at the option we ultimately choose compared to one we ultimately do not choose (Glaholt & Reingold, 2009a, 2009b, 2011; Glaholt, Wu & Reingold, 2009; Pieters & Warlop, 1999; Schotter et al., 2010; Shimojo, Simion, Shimojo & Scheier, 2003; Simion & Shimojo, 2006). Furthermore, this gaze bias is manifested as both an increasing probability of looking at the chosen item in the time leading up to the choice (Shimojo et al., 2003; Simion & Shimojo, 2006) and increased frequency of viewing (*dwell frequency*) and amount of time (*dwell time*) on the ultimately chosen option even during the first encounter (Glaholt & Reingold, 2009a, 2009b; Glaholt et al., 2009; Schotter et al., 2010). Thus, it seems as if subjects may initially know that they might be more likely to choose an item, even if they do not know what the other options are. These findings suggest that, during initial encoding, subjects are not passively sampling various options, but actively evaluating them, online.

To explain this latter finding, Glaholt and Reingold (2009a, 2009b, 2011) and Schotter et al. (2010) suggested that gaze bias may be a more general aspect of the decision-making process and that the early gaze bias effect may be related to selective encoding. Mainly, people spend longer entertaining an option that may potentially satisfy the choice they are trying to make and exclude an option, relatively quickly, if it shows little potential to be chosen. More specifically, Schotter et al. (2010) showed that the *gaze bias effect* –longer time looking at the chosen compared to not chosen option—arises in early encounters with the option. That is, *first dwell time* (the amount of time spent looking at an option during the first encounter before leaving it) was modulated by some manipulations, like decision prompt (what subjects are told to base their decision on), but not others. In contrast, the gaze bias effect in *remaining time* (any time spent on an option after the first dwell) was not sensitive to such manipulations.

Schotter et al. had subjects make a two-alternative forced choice decision between two images with four different decision prompts. The decision prompts were two opposites of a dichotomy: like/dislike, "Which one of these images do you like more/less?" and older/newer, "Which one of these images do you think is older/was taken more recently?" In first dwell time, they found an interaction between the gaze bias effect and decision prompt for the like/dislike conditions (i.e., a large gaze bias effect for the like decision but no gaze bias effect for the dislike decision), but no interaction for the older/newer decisions (i.e., an equivalent, moderately sized gaze bias effect for the older and newer decisions).

Schotter et al. suggested that the interaction for the like/dislike decisions arose from the contribution of two biases—one to spend more time on the option they like (a *liking effect*) and one to spend more time on the option that best satisfies the criterion of the decision prompt (the general *gaze bias effect*—reflecting selective encoding). The exaggerated difference in first dwell time between the chosen and not chosen item in the like decision condition and null difference in the dislike decision condition suggests that the two biases that can operate either additively (in the like decision condition) or subtractively (in the dislike decision condition). In contrast, the lack of an interaction for the difference in dwell times in the older/newer decision conditions suggests that there may only be one bias (the *gaze bias*) involved in those decision tasks. Schotter et al. found no interactions between decision prompt and the gaze bias effect in remaining time and suggested that biases may only be malleable (i.e., show interactions) during the encoding stage of the decision-making process, not the post-encoding, selection stage.

Schotter et al. suggested that selective encoding (see also Glaholt & Reingold, 2009a, 2009b, 2011) is what drives the differences in the patterns of data they reported. Mainly, the gaze bias effect in remaining time should be relatively constant across manipulations, because it reflects a time at which options have already been encoded and simultaneously evaluated and all that needs to be done is selection. In contrast, according to the selective encoding idea, interactions should arise for the gaze bias effect in first dwell time only if that factor is relevant for the decision. Therefore, gaze bias should not interact with decision prompt if the prompts are equivalent and there are no existing biases that would lead subjects to look longer at one option versus the other (e.g., the older/newer decision). However, the gaze bias effect should be modulated by a factor that is relevant or when there are existing biases that either support or contradict the decision to be made (e.g., the like/dislike decision).

More generally, selective encoding suggests that these interactions reflect an active, online evaluation of options during the first encounter. Furthermore, which properties of the stimulus cause modulations of the gaze bias effect should be those that are task-relevant and useful as a heuristic to make the choice at hand. Thus, subjects should spend longer looking at an option that satisfies the decision criterion (i.e., how much it coincides with the heuristic). Importantly, this is only true of the first encounter with the stimulus. In contrast, the remaining time spent during the decision process should be less sensitive to these heuristics and should, therefore, be robust to such manipulations.

To test if this is the case, we sought to directly manipulate a variable that may influence the gaze bias effect during the encoding phase of the older/newer decision tasks: the usefulness of a color content cue to indicate recentness. Because in the Schotter et al. (2010) study, older/newer decision prompt did not interact with the gaze bias effect we needed to find an image property that might be easily encoded and used as a heuristic to make such decisions. We chose color content: whether the image was displayed in color or in black-and-white because data from the Schotter et al. (2010) study suggest that color content may be a selectively encoded stimulus property for an older/newer decision. However, while Schotter et al. did vary whether the pair of images were both color or both black-and-white, they did not manipulate the color content of the images directly and did not know which image was older than the other, and thus could not test this hypothesis directly. In the present study we chose pairs of images for which one was clearly older, based on the semantic content. Furthermore, we orthogonally manipulated whether an image was displayed in color or black-and-white. Additionally, we manipulated, across blocks of the experiment whether, color content was a useful heuristic for determining the relative oldness of the images. If the early gaze bias effect reflects selective encoding of the stimuli we should see an interaction in the magnitude of the gaze bias effect across blocks during first dwell time, but not remaining time. Furthermore, this interaction should be present when color content differs between the pair of images (and would have the potential to be used as a heuristic) and absent when it is the same.

Method

Subjects

Thirty-two undergraduates from the University of California, San Diego participated in the experiment. They received course credit in exchange for participation. All subjects had normal or corrected-to-normal vision

$watermark-text
$watermark-text

Apparatus

Eye movements were recorded via a SR Research Ltd Eyelink 1000 eyetracker. Viewing was binocular but only movements from the right eye were recorded. Following calibration, eye position errors were less than 1.0°. Stimuli were presented on a 20 inch Sony Trinitron CRT monitor with a pixel resolution of 1280 × 1024 and subjects were seated approximately 60 cm away from the monitor.

Materials and Design

The stimuli consisted of 100 pairs of color photographs (200 pictures in total), obtained from the Internet. The photos varied in subject matter (landscapes, street scenes, portraits, etc.) but pairs of images were selected so that they shared the same content (e.g. both portraits, both street scenes, etc.). Of the pairs of images, one was clearly taken more recently, given the content of the image. For example, for a pair of images that contained planes, one image contained a picture of a World War II plane and one contained a picture of a Boeing 737. Each image was manipulated by removing the color information to yield a black-and-white version; color versions and black-and-white versions were used in the experiment.

The black-and-white and color versions of each pair of images were combined to create four different conditions per pair: (1) both black-and-white, (2) both-color, (3) the older image was black-and-white and the newer image was color, and (4) the older image was color and the newer image was black-and-white.

The experiment consisted of 4 blocks of trials. In half of the blocks the subject was asked to determine which image was taken longer ago (older decision). In the other blocks the subject was asked to determine which image was taken more recently (newer decision). Orthogonal to the decision prompt manipulation, we manipulated the consistency of color content as an indication of recentness of the photo.

In the consistent blocks, when subjects were making an older decision, the pair of images were either both presented in black-and-white or the older image was black-and-white and the newer image was color. Conversely, when subjects were making a newer decision, the pair of images were either both color or the newer image was color and the older image was black-and-white. Thus, the correct image always matched what would be predicted by color content, but when the images contained the same color contents, this could not be used as a heuristic to accurately select the correct image.

In the inconsistent blocks, when the color content was the same between images they were both color half the time and both black-and-white half the time. Additionally, when color content differed, the correct image matched the color content only half the time (e.g., the newer image was color and the older image was black-and-white on half the trials).

The experiment was fully counterbalanced in a Latin-square design so that all pairs of photographs appeared in all conditions, across subjects and all subjects saw an equal number of pairs in each condition. However, the two consistent blocks were always seen first and the two inconsistent blocks were always seen second. This was done to establish how much subjects relied on color content as a heuristic when it was completely reliable. Presenting the inconsistent block second would then allow us to see how much they discounted color content when it was no longer reliable. The order of older decision and newer decision within consistent and inconsistent blocks was counterbalanced across subjects.

Picture pairs were presented on a white background with one photograph on the left side of the screen and one on the right. Each picture was the same width (14.69 degrees of visual angle) and varied slightly in height.

Procedure

At the beginning of the experiment the experimenter informed the subject that he/she was going to be making judgments about pairs of pictures. Before each block the experimenter told the subject what decision to make (i.e., older or newer) and told them to make that decision for every pair of pictures in the block. The eyetracker was calibrated before each block to ensure that the eye movement data was as accurate as possible. Once the calibration was successful, the experiment began. At the beginning of each trial a fixation point was presented at the center of the screen. Once the subject fixated this point the experimenter pressed a button to make the pair of pictures appear. The subject made a decision by pressing a button on a video game controller corresponding to their choice: the left button to choose the left picture and the right button to choose the right picture. Once they made their decision the fixation point appeared to start the next trial.

Results and Discussion

Trials were separated into single dwell trials (i.e., those in which there were, at maximum, one dwell on each image; 40% of the data) and multiple dwell trials (i.e., those in which there were at least two dwells on at least one of the images; 60% of the data)[1]. We separated these trials because they might reflect qualitatively different decision processes (i.e., those in which the decision can be made easily, with no more than one dwell—single dwell trials—and those in which the decision is more difficult and more information must be gathered—multiple dwell trials). For single dwell trials, we excluded trials if subjects fixated either image for between 1 and 79 ms (to retain information about completely un-fixated items; 1% of the data). For multiple dwell trials, we excluded trials in which subjects did not fixate either image for at least 80 ms (<1% of the data). In the reported analyses we did not exclude trials in which subjects responded incorrectly (only 8% of the data in single dwell trials and 10% of the data in multiple dwell trials) so that the reported data from the accuracy analyses and the eye movement analyses would reflect the same trials. Furthermore, subjects did not know whether their responses were correct or incorrect, so it's likely that the eye movement behavior they exhibited in both correct and incorrect trials reflect the same decision process. Analyses with incorrect trials excluded showed the same pattern of data as those with these trials included.

We measured response accuracy as well as the same looking times reported by Schotter et al. (2010): *first dwell time* (all time spent fixating the image before first leaving it), *remaining time* (all time spent looking at the image after the first dwell), and *dwell frequency* (how many separate times the image was viewed). To analyze the data we ran linear mixed effects models with crossed random effects of subjects and items (Baayen, Davidson, & Bates, 2008), which are similar to ANOVAs but more robust to uneven data (produced by splitting the data between single and multiple dwell trials) and more amenable to testing specific contrasts with different number of observations (as there are here). To analyze binary data (i.e., all data reflecting the subject's choices, rather than viewing behavior) we ran logistic regressions (linear mixed effects models using a logit transform to fit binary data; Jaeger, 2008). For binary data we report z statistics and p values. For continuous data (i.e., eye movement measures) we report regression coefficients (b), which estimate the effect size, of the reported comparison, standard error, and the t value of the effect coefficient but not the t value's degree of freedom because there is no consensus in the literature for what degrees of freedom to use. The p-values were estimated using Markov-chain Monte Carlo sampling.

[1]We thank Eyal Reingold for this suggestion.

Accuracy

The proportion of correct responses as a function of block and color match (separated for single dwell and multiple dwell trials) as well as number of trials are shown in Table 1. Logistic regressions revealed very similar results between single dwell and multiple dwell trials, although subjects were slightly more accurate in single dwell trials (indicating that there was not a speed-accuracy tradeoff). The logistic regressions show that subjects were more accurate in the consistent block (96% in single dwell trials and 94% in multiple dwell trials) than in the inconsistent block (89% in single dwell trials; z = 4.24, $p < .001$ and 87% in multiple dwell trials; z = 5.07, $p <.001$) suggesting that reliability of the color content cue improved accuracy. Additionally, there was a significant effect of color match (single dwell trials: z = 2.37, $p < .05$; multiple dwell trials: z = 3.58, $p < .001$) such that subjects were more accurate when the images had different color contents (i.e., when one was black and white and the other was color; 94% in single dwell trials and 92% in multiple dwell trials) than when they had the same color content (i.e., when they were both black and white or when they were both color; 91% in single dwell trials and 89% in multiple dwell trials). There was a significant interaction in multiple dwell trials between block and color match (z = 3.38, $p < .001$) such that the effect of color match was greater in the consistent block (a 7% difference) than in the inconsistent block (a <1% difference). The interaction was not significant in single dwell trials ($p > .1$).

These data suggest that subjects relied, to some extent on the color content of the images to make their decision. Therefore, when the color content was the same (and color content could not be used as a heuristic), subjects were less accurate than when they were different (and color content could be used as a heuristic). Furthermore, the extent to which the color content heuristic was reliable significantly influenced the extent to which subjects performed accurately. Thus, in the inconsistent block, where color content was not reliable (i.e., when the older image was black-and-white and the newer image was color only half the time) subjects' accuracy significantly decreased (from 97% to 90% for single dwell trials and from 97% to 88% in multiple dwell trials). However, across the two blocks, trials in which the color content matched (i.e., they were both black and white or they were both color) subjects' accuracies did not change much (from 93% to 88% in single dwell trials and from 90% to 87% in multiple dwell trials). These data suggest that subjects did use the color content of the images to make their decisions. Thus, when this was a useful heuristic (in the consistent block) subjects performed significantly better than when this was not a useful heuristic (in the inconsistent block). Lastly, the pattern results were very similar between single dwell and multiple dwell trials, indicating that the decision that subjects ultimately came to did not differ between different types of trials (e.g., those in which they needed to look at the images once or less or those in which they needed to look at them multiple times).

Eye movement measures

As mentioned above, we analyzed *first dwell time* (Figure 1), *dwell frequency* (Figure 2), and *remaining time* (for multiple dwell trials only because remaining time can only be calculated when subjects viewed at least one image more than once; Figure 3). Means and standard errors for single dwell trials are shown in Table 2 and means and standard errors for multiple dwell trials are shown in Table 3. We will report analyses for single dwell trials and multiple dwell trials separately because the represent qualitatively different decision processes.

Single Dwell Trials

First Dwell Time—Linear mixed effects models revealed a significant effect of block (b = 28.47, SE = 10.55, $t = 2.70$, $p < .01$), indicating that subjects spent longer looking at the images in the inconsistent block (M = 523 ms) compared to the consistent block (M = 520 ms). There was a significant effect of color match (b = 31.87, SE = 10.49, $t = 3.04$, $p < .005$), indicating that subjects spent longer on the pictures when the color content was the same (M = 529 ms) than when it was different (M = 514 ms). There was a significant gaze bias effect (b = 139.23, SE = 10.20, $t = 13.65$, $p < .001$) with more time spent looking at the chosen picture (M = 599 ms) than the not chosen picture(M = 444 ms).

There was a significant interaction between block and color match (b = -63.30, SE = 20.99, $t = 3.02$, $p < .005$) indicating that the effect of color matching between the two images had a very different effect in the consistent block (where subjects spent longer on images when the color content was the same (543 ms) than when it was different (498 ms)) compared to the inconsistent block (where subjects spent longer on images when color content was different (530 ms) than when it was the same (516 ms)). This interaction suggests that subjects were sensitive to the reliability of the color content cue. In the block in which it was always reliable, they used the cue to make their decision and were faster to view the images than when the color content was the same. In contrast, in the inconsistent block, because color content was misleading half of the time, subjects spent more time viewing the images when color content was different, because they needed more time to determine whether they should use the cue or not. The interaction between the gaze bias effect and block was marginally significant (b = -37.55, SE = 20.41, $t = 1.84$, $p < .10$) with the gaze bias effect being larger in the consistent block (a 164 ms difference) than in the inconsistent block (a 144 ms difference). These data also suggest that subjects were sensitive to the color content heuristic in that the degree to which they could exclude an item from consideration during the encoding phase was modulated by the reliability of the color content cue. All other interactions were not significant (both ps > .1). Taken together, these data indicate that, in trials in which subjects were able to make a decision by viewing the images once or fewer times, how long they spent looking at the images in order to make such a decision was modulated by their ability to use a task-relevant heuristic: color content.

Dwell Frequency—Because single dwell trials are those in which subjects looked at an image at minimum zero times and at maximum one time (i.e., the data represent a binary outcome), we used a logistic regression to analyze dwell frequency. The logistic regression revealed no significant effects or interactions (all *p*s > .98). This is not surprising, given that in the majority of the time in single dwell trials (97%) subjects viewed each image exactly one time. This lack of variability leaves little opportunity for any factor to significantly affect dwell frequency. Given that there is very little variability in dwell frequency for single dwell trials we will not analyze or discuss these data further.

Multiple dwell trials

First Dwell Time—Linear mixed effects models revealed no significant effect of block (b = -2.55, SE = 9.51, $t < 1$), indicating that subjects did not spend longer looking at the images during their first dwell in the inconsistent block compared to the consistent block (both Ms = 527 ms). Despite the null effect of block, there was a significant effect of color match (b = 23.30, SE = 9.50, $t = 2.45$, $p < .05$), indicating that subjects spent longer on the pictures when the color content was the same (M = 541 ms) than when it was different (M = 512 ms). There was a significant gaze bias effect (b = 23.02, SE = 9.33, $t = 2.47$, $p < .05$) with more time spent looking at the chosen picture (M = 538 ms) than the not chosen picture (M = 515 ms).

There was a significant interaction between the gaze bias effect and color match (b = -39.95, SE = 18.70, $t = 2.14$, $p < .05$) indicating that the gaze bias effect was larger when the color content of the two images was different (a 40 ms effect) than when the color content was the same (a 7 ms effect). These data suggest that, as seen above in the single dwell trials, during subjects' first encounter with the images, viewing time was significantly affected by the reliability of the task-relevant heuristic: color content. All other interactions were not significant (both *p*s > .25).

Dwell Frequency—Because the number of views in multiple dwell trials ranged from one view to six views we were able to use linear mixed effect models to analyze the data (as opposed to the logistic regressions used in single dwell trials). Linear mixed effect models revealed that the effect of block was significant (b = .05, SE = .02, t = 2.32, $p < .05$), indicating that subjects viewed the images more times in the inconsistent block (M = 1.81 times) than in the consistent block (M = 1.76 times). There was also a significant effect of color match (b = 0.06, SE = 0.02, t = 3.04, $p < .005$) indicating that subjects viewed the images more times when the color content was the same (M = 1.82 times) than when it was different (M = 1.75 times). There was a significant gaze bias effect (b = 0.26, SE = 0.02, t = 12.34, $p < .001$) indicating that subjects looked more times at the chosen image (M = 1.9 times) than the not chosen image (M = 1.66 times).

There was a significant interaction between color match and block (b = -0.10, SE = .04, t = 2.29, $p < .05$) with an effect of color content matching in the consistent block (a difference of .15 views) but not in the inconsistent block (both Ms = 1.81). These data indicate that, when subjects could use color content to distinguish and compare the images, they could make their decision faster (i.e., they did not need to view the images as many times to make their decision). None of the other interactions were significant (all *p*s > .15). Remaining Time. Linear mixed effects models revealed a significant effect of block (b = 73.01, SE = 20.89, $t = 3.50$, $p < .005$), indicating that subjects spent longer looking at the images during the remaining time in the trial in the inconsistent block (M = 446 ms) compared to the consistent block (M = 378 ms). There was a significant effect of color match (b = 77.71, SE = 20.88, $t = 3.72$, $p < .001$), indicating that subjects spent longer when the color content was the same (M = 456 ms) than when it was different (M = 369 ms). There was also a significant gaze bias effect (b = 100.19, SE = 20.48, $t = 4.89$, $p < .001$) with more time spent looking at the chosen image (M = 454 ms) than the not chosen image (M = 371 ms).

As expected, there were no significant interactions in remaining time (all ps > .25) because remaining time represents the post-encoding phase when subjects are only selecting the already-encoded items and are no longer sensitive to the color content heuristic.

To summarize the above data, during first dwell in both single dwell trials and multiple dwell trials, viewing times were modulated by the reliability of the color content heuristic. In contrast, as expected, in remaining time in multiple dwell trials, when the stimuli had already been encoded and evaluated, the gaze bias effect was not modulated by the reliability of the color content heuristic.

Color content reliability for color different trials

To directly assess the effect of reliability of the color content heuristic on a trial to trial basis we compared the gaze bias effect across trials where color content differed between the pair of images. Note that, in the consistent block, when color content differed between the two images, the color content heuristic was always reliable, and therefore useful for the decision task. In contrasts, in the inconsistent block, the heuristic was only useful half of the time. Therefore, on a given trial, it was unclear, at first, whether color content should be used to make the decision, when color content differed between the two images. If subjects are sensitive to the reliability of the cue, in general, we would expect to see an effect of block (consistent vs. inconsistent) such that subjects respond to. If subjects were sensitive to the reliability of color content on a trial-to-trial basis we would expect to see a difference between reliable and not reliable trials within the inconsistent block. To test these predictions, we created a three-level factor that coded trials as either (a) in the consistent block, (b) in reliable trials (i.e., those in which the older image was black-and-white) in the inconsistent block, or (c) in unreliable trials (i.e., those in which the older image was color) in the inconsistent block. In the linear mixed effects models we set specific contrasts to test (1) whether there was an overall effect of block (consistent block vs. inconsistent block; -1, .5, .5) and (2) whether, within the inconsistent block, there was a difference between trials in which the color content was reliable and when it was not reliable (0, -1, 1). As mentioned above, there is no remaining time in single dwell trials and very little variability in dwell frequency in single dwell trials, so we only report first dwell time for single dwell trials.

Single Dwell Trials

The analysis of first dwell time revealed a main effect of block (b = 57.54, SE = 16.35, $t = 3.52$, $p < .001$) indicating that subjects spent longer in the inconsistent block (M = 525 ms) than the consistent block (M = 498 ms). Within the inconsistent block, however, the contrast between reliable trials (M =538 ms) and not reliable trials (M = 513 ms) was not significant ($p > .25$). There was a significant gaze bias effect (b = 145.23, SE = 16.87, $t = 8.61$, $p < .001$) indicating longer time looking at the chosen (M = 590 ms) than the not chosen image (M = 441 ms).

The interaction between the gaze bias effect and block was marginally significant (b = -52.02, SE = 31.39, $t = 1.66$, $p < .10$) with the gaze bias effect being larger in the consistent block (a 162 ms difference) than in the inconsistent block (a 139 ms difference). These data indicate that subjects were generally sensitive to the reliability of color content to indicate recentness. However, the interaction between the gaze bias effect and reliability within the inconsistent block was not significant ($t < 1$), indicating that the gaze bias effect was not affected by the reliability of the color content heuristic on each trial. Taken together, these data suggest that subjects became more reluctant to rely on the color content heuristic in general in the inconsistent block, but were not able to distinguish, in the first dwell, whether the color content on the individual trial was reliable or not.

Multiple Dwell Trials

First Dwell Time—The analysis revealed that the effect of block was not significant ($t < 1$) indicating that subjects did not spend significantly longer in the inconsistent block (M = 518 ms) than the consistent block (M = 508 ms). Additionally, within the inconsistent block, the contrast between reliable trials (M = 508 ms) and not reliable trials (M = 528 ms) was not significant ($t < 1$). There was a significant gaze bias effect (b = 40.16, SE = 14.06, $t = 2.86$, $p < .005$) indicating longer time looking at the chosen (M = 535 ms) than the not chosen image (M = 495 ms).

Neither the interaction between the gaze bias effect and block nor the interaction between the gaze bias effect and reliability within the inconsistent block were significant (both *p*s > .25). These data indicate that, in trials in which color content differed between the images and the decision was difficult (i.e., subjects needed to view the images multiple times), the gaze bias effect in first dwell was not affected by the reliability of the color content heuristic.

Remaining time

In remaining time, there was a significant main effect of block (b = 98.76, SE = 29.34, $t = 3.37$, $p < .001$), indicating that subjects spent longer in the inconsistent block (M = 431 ms) than in the consistent block (M = 307 ms). However, within the inconsistent block, the difference between reliable trials (M = 424 ms) and not reliable trials (M = 439 ms) was not significant ($t < 1$). There was a significant gaze bias effect (b = 100.35, SE = 28.92, $t = 3.47$, p < .001) with longer looking times on the chosen (M = 431 ms) than the not chosen image (M = 349 ms), but this did not interact with block ($t < 1$) or reliability within the inconsistent block ($p > .15$). Again, as expected, in remaining time (once the information had already been encoded and evaluated), the gaze bias effect was not modulated by the task-relevant heuristic.

Dwell frequency—There was a main effect of block (b = .11, SE = .03, $t = 3.56$, $p < .001$) with more views of the images in the inconsistent block (M = 1.81 views) than the consistent block (M = 1.68 views). Within the inconsistent block, the difference between reliable trials (M = 1.80 views) and not reliable trials (M = 1.83 views) was not significant ($t < 1$). There was a significant gaze bias effect (b = .22, SE = .03, $t = 7.25$, $p < .001$) with more views on the chosen image (M = 1.87 views) than the not chosen image (M = 1.66 views).

The interaction between the gaze bias effect and block was not significant ($t < 1$). However, the effect of color content reliability within the inconsistent block was marginally significant (b = -0.14, SE = .08, $t = 1.74$, $p < .10$), indicating that subjects looked at the images a similar number of times across blocks, but within the inconsistent block, they needed to view the images more times when the color content cue was unreliable.

Color content reliability for color same trials

In addition to assessing reliability of the color content heuristic in color different trials, we performed similar analyses on color same trials. Note that in the consistent block, the images were always both black-and-white for older decisions and both color for newer decisions. However this was only true half the time in the inconsistent block. We do not expect that there should be significant effects of block or reliability because the color content heuristic (difference in color content between the two images) is absent in these trials and it is therefore unlikely that subjects would behave differently in these trials between blocks or between different trials within the inconsistent block. As in the linear mixed effects models reported above, we set specific contrasts to test (1) whether there was an overall effect of block (consistent block vs. inconsistent block) and (2) whether, within the inconsistent block, there was a difference between trials in which the color content matched the decision (i.e., when the pair was color in a newer decision or the pair was black-and-white in an older decision).

Single Dwell Trials

There was a significant gaze bias effect with longer looking times on the chosen item (M = 601 ms) than the not chosen item (M= 453 ms; b = 118.27, SE = 13.12, $t = 9.02$, $p < .001$). Neither the effect of block, nor the effect of reliability within block, nor any of the interactions were significant (all *t*s < 1).

Multiple Dwell Trials

Across all measures, none of the effects were significant (all *p*s > .1) except for the gaze bias effect in remaining time (M_{chosen} = 514 ms, $M_{not\ chosen}$ = 409 ms; b = 111.56, SE = 30.79, $t = 3.62$, $p < .001$) and dwell frequency (M_{chosen} = 1.96 views, $M_{not\ chosen}$ = 1.69 views; b = .29, SE = .03, $t = 9.29$, $p < .001$).

These data suggest that subjects only use color content as a heuristic when color content differs between the images and, therefore, reliability of the color content does not have an effect on the decision process.

Conclusion

Our data show that subjects reliably exhibit a gaze bias effect in many different stages of the decision-making process: first dwell time (the first or only time the image is encountered and encoded), remaining time (the post-encoding decision stage when there are multiple dwells on an image), and dwell frequency (the number of encounters required to make a decision when there are multiple dwells on an image). Crucially, however, as predicted, the magnitude of the gaze bias effect was modulated by a decision-related heuristic (color content reliability) only in first dwell time (when the images are being encoded) for both single dwell and multiple dwell trials. There were no interactions in remaining time in any of the analyses. These data suggest that, when making a decision, people are sensitive to properties of the options that are relevant to the decision at hand (Glaholt & Reingold, 2009a, 2009b, 2011; Schotter et al., 2010). Furthermore, these data suggest that people can modulate the extent to which they are sensitive to those relevant features of an option, depending on how reliable that feature is in indicating the correct answer. Importantly, the stage in the decision process at which subjects are flexibly sensitive to such heuristics is the encoding and evaluation stage. Thus, it seems as if, after the options have been encoded and evaluated, these heuristics are no longer relied on to make a decision.

Additionally, the only interaction with the gaze bias effect that was observed in single dwell trials involved block, not reliability of the color content cue within the inconsistent block. This suggests a general sensitivity to heuristic reliability (i.e., subjects are sensitive to the fact that, in the inconsistent block, the heuristic is no longer useful) and not an on-line sensitivity to whether the color content on an individual trial is useful. This is not surprising, since subjects cannot know a priori whether the cue will be reliable on a given trial. Thus, it would seem prudent to stop relying on the cue altogether in a block in which it is unclear whether it should be used or not. Furthermore, there were no effects of or interactions with block in trials in which color content was the same between the images. This is expected, given that subjects would be unlikely to use the color content heuristic in situations in which it does not distinguish the images.

Together, with the results of Schotter et al. (2010) these data suggest that people are highly adaptive with respect to how they gather information in order to make a decision. Schotter et al. (2010) reported an interaction between the gaze bias effect and decision prompt (like vs. dislike decisions) as well as an interaction between the gaze bias effect and color content matching between the images. The present study reports an interaction between the gaze bias effect and color content matching as well as an interaction between the gaze bias effect and color content reliability (consistent vs. inconsistent blocks). Most importantly for the idea of selective encoding, in both experiments, these interactions only arise in first dwell time (the encoding and evaluation stage), not remaining time (the post-encoding, selection stage). These data suggest that, in contrast to some models of decision-making (Busmeyer & Townsend, 1993; Johnson & Busemeyer, 2005), the information gathering stage and evaluation stages are not separate processes, but rather occur concurrently and dynamically when the options are first encountered.

Finally, these data, along with many other studies (Glaholt & Reingold, 2009a, 2009b; Glaholt, Wu & Reingold, 2009; Pieters & Warlop, 1999; Schotter et al., 2010; Shimojo, Simion, Shimojo & Scheier, 2003; Simion & Shimojo, 2006; see Glaholt & Reingold, 2011 for a review) emphasize the usefulness and importance of using eye tracking methodology to study decision-making processes. Eye tracking provides rich information about both where and when subjects are attending in real time (Rayner, 1998, 2009) and can therefore shed light on various time points and substages within complex processes, such as decision-making.

Acknowledgments

This research was supported by a Pre-doctoral Fellowship on Training Grant DC000041 from the National Institute on Deafness and Other Communication Disorders awarded to the first author and Grants HD065829 and HD26765 from the National Institutes of Child Health and Human Development. We would like to thank Raymond W. Berry for help developing the stimuli for this project and Alex Kadokura for help running subjects.

References

Baayen R, Davidson D, Bates D. Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language. 2008; 59:390–412.

Busemeyer JR, Townsend JT. Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. Psychological Review. 1993; 100:432–459. [PubMed: 8356185]

Glaholt MG, Reingold EM. The time course of gaze bias in visual decision tasks. Visual Cognition. 2009a; 17:1228–1243.

Glaholt MG, Reingold EM. Stimulus exposure and gaze bias: a further test of the gaze cascade model. Attention, Perception & Psychophysics. 2009b; 71:445–450.

Glaholt MG, Reingold EM. Eye movement monitoring as a process tracing methodology in decision making research. Journal of Neuroscience, Psychology, and Economics. 2011; 4:125–146.

Jaeger TF. Categorical data analysis: Away from ANOVAs (transformation or not) and toward logit mixed models. Journal of Memory and Language. 2008; 59:434–446. [PubMed: 19884961]

Johnson JG, Busemeyer JR. A dynamic, stochastic, computational model of preference reversal phenomena. Psychological Review. 2005; 112:841–861. [PubMed: 16262470]

Pieters R, Warlop L. Visual attention during brand choice: the impact of time pressure and task motivation. International Journal of Research in Marketing. 1999; 16:1–16.

Rayner K. Eye movements in reading and information processing: 20 years of research. Psychological Bulletin. 1998; 124:372–442. [PubMed: 9849112]

Rayner K. The Thirty-fifth Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. The Quarterly Journal of Experimental Psychology. 2009; 62:1457–1506. [PubMed: 19449261]

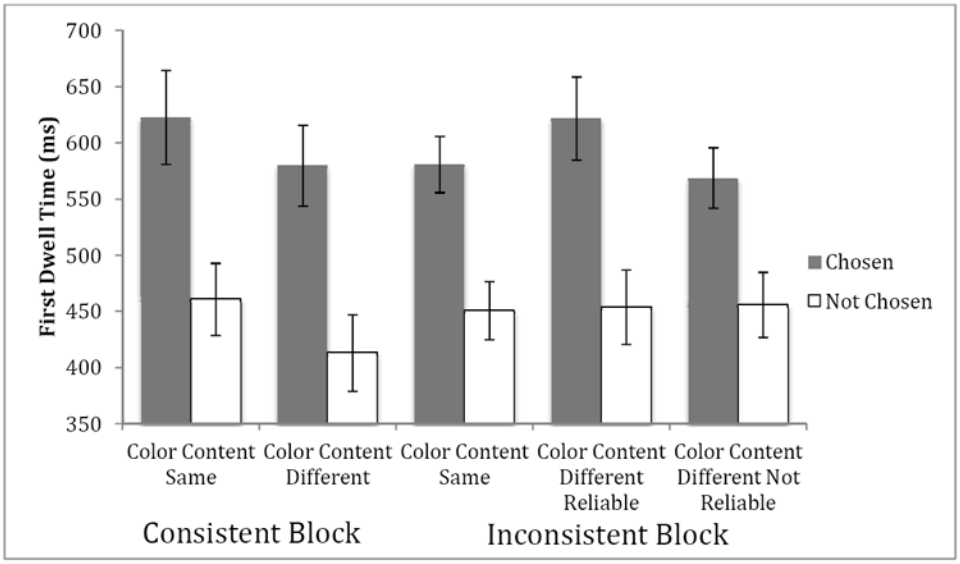
Schotter ER, Berry R, McKenzie C, Rayner K. Gaze bias: Selective encoding and liking effects. Visual Cognition. 2010; 18:1113–1132.

Shimojo S, Simion C, Shimojo E, Scheier C. Gaze bias both reflects and influences preference. Nature Neuroscience. 2003; 6:1317–1322.

Simion C, Shimojo S. Early interactions between orienting, visual sampling and decision making in facial preference. Vision Research. 2006; 46:3331–3335. [PubMed: 16765404]

Simon C, Shimojo S. Interrupting the cascade: orienting contributes to decision making even in the absence of visual stimulation. Perception & Psychophysics. 2007; 69:591–595. [PubMed: 17727112]

A



B

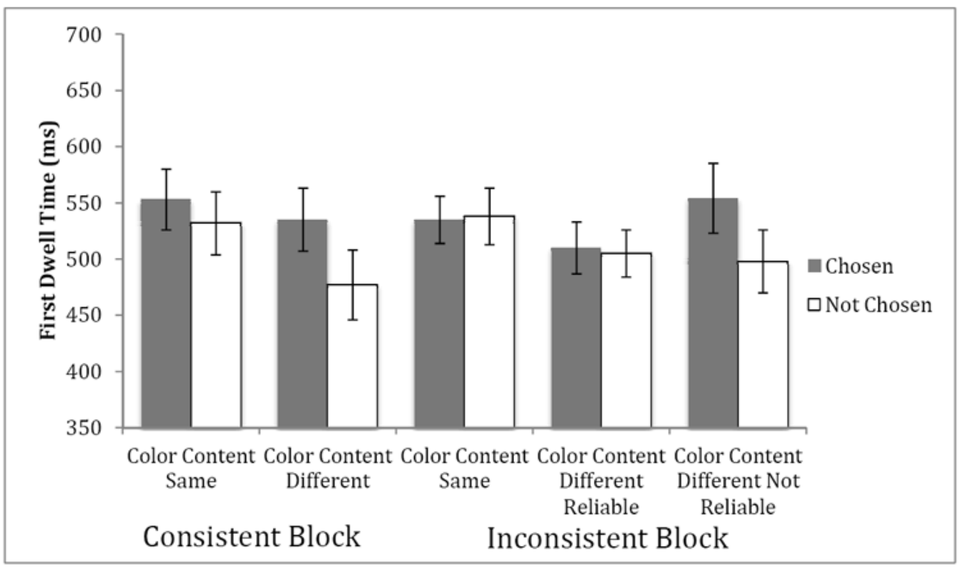


Figure 1.

First dwell time on the chosen and not chosen image as a function of block and color content. Error bars represent standard error of the mean. Panel A represents single dwell trials. Panel B represents multiple dwell trials.

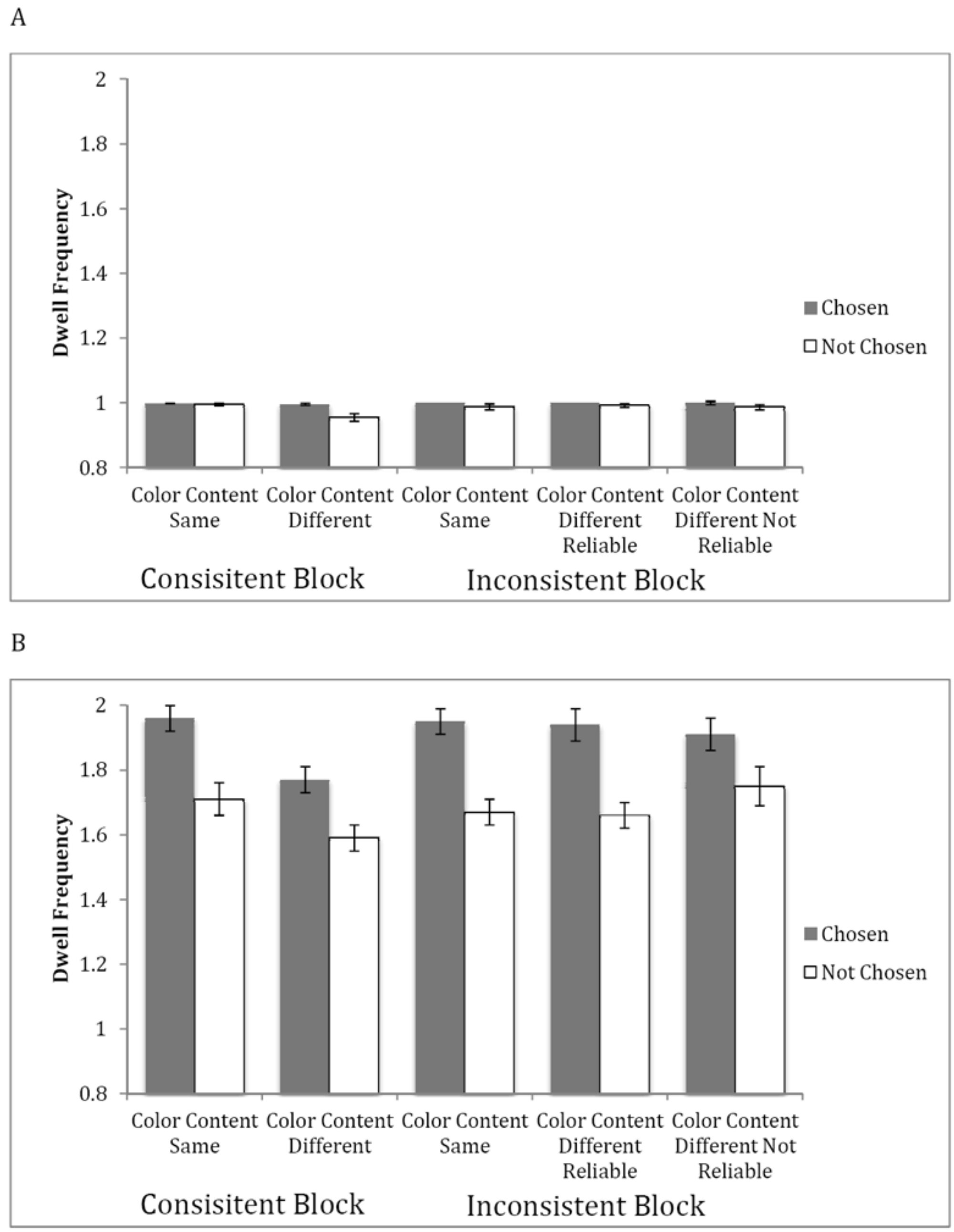


Figure 2.
Dwell frequency on the chosen and not chosen image as a function of block and color content. Error bars represent standard error of the mean. Panel A represents single dwell trials. Panel B represents multiple dwell trials.

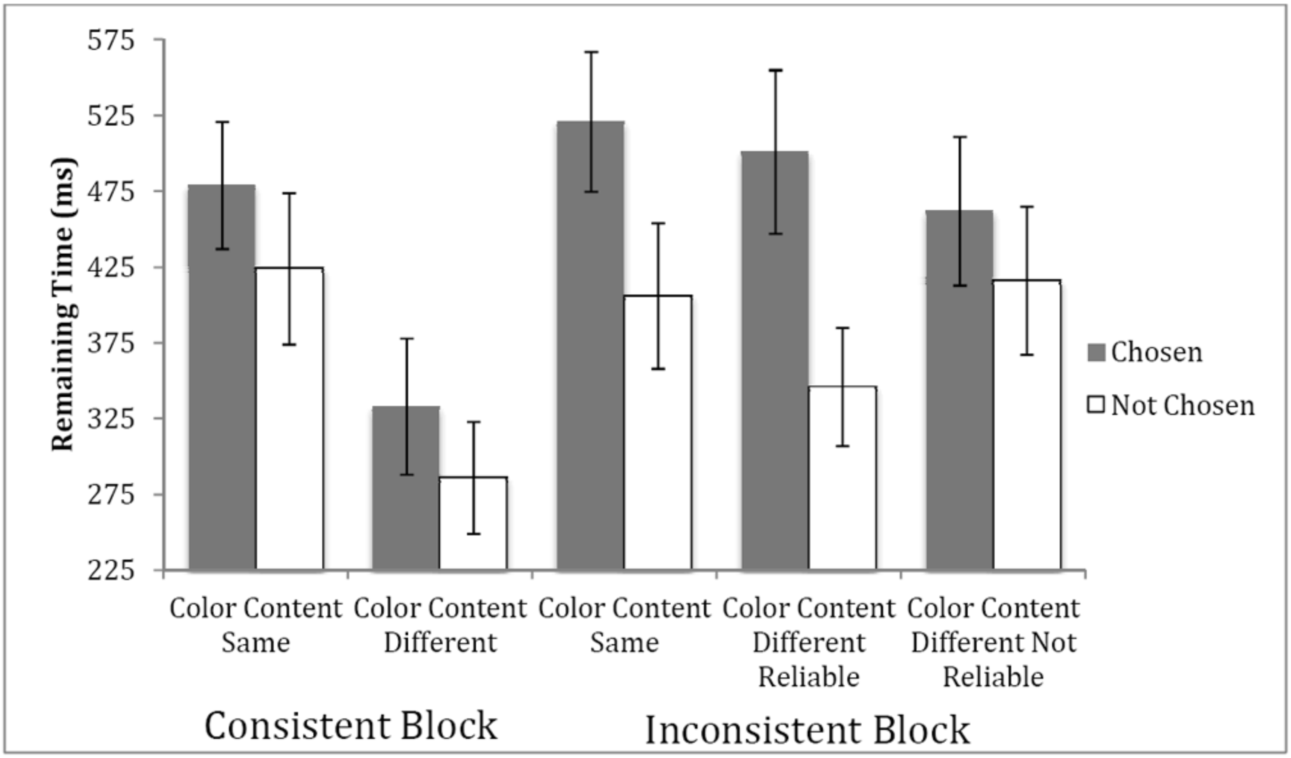


Figure 3.
Remaining time on the chosen and not chosen image as a function of block and color content in multiple dwell trials. Error bars represent standard error of the mean.

Table 1

Mean accuracy and number of trials as a function of block and color content for single dwell trials and multiple dwell trials.

	Block		Color content different		Color content same	
			Accuracy	N	Accuracy	N
Single Dwell Trials						
	Consistent		98%	359	93%	283
	Inconsistent	Reliable	87%	143	88%	297
		Not reliable	93%	174		
Multiple Dwell Trials						
	Consistent		97%	418	90%	514
	Inconsistent	Reliable	86%	257	87%	497
		Not reliable	89%	222		

Table 2

Means and standard errors of data from single dwell trials for first dwell time and dwell frequency as a function of block, color content and whether the image was chosen or not chosen.

Block	Color Content	Image	First Dwell Time	Gaze Bias Effect	Dwell Frequency
Consistent	Same	Chosen	624(42)	162	.999 (0.001)
		Not Chosen	462(32)		.994 (0.004)
	Different	Chosen	582(36)	167	.996 (0.003)
		Not Chosen	415(34)		.951 (0.013)
Inconsistent	Same	Chosen	581 (25)	130	1.00 (0.000)
		Not Chosen	451 (26)		.988 (0.009)
	Different Reliable	Chosen	622 (37)	168	1.00 (0.00)
		Not Chosen	454 (33)		.992 (0.005)
	Different Not reliable	Chosen	569 (30)	112	1.00 (0.005)
		Not Chosen	457(29)		.986 (0.008)

Table 3

Means and standard errors of data from multiple dwell trials for first dwell time, remaining time and dwell frequency as a function of block, color content and whether the image was chosen or not chosen.

Block	Color Content	Image	First Dwell Time	Gaze Bias Effect	Remaining Time	Dwell Frequency
Consistent	Same	Chosen	553 (26)	17	479 (42)	1.96 (0.04)
		Not Chosen	536(28)		424 (50)	1.71 (0.05)
	Different	Chosen	538 (28)	59	333 (45)	1.77 (0.04)
		Not Chosen	479 (31)		286 (37)	1.59 (0.04)
Inconsistent	Same	Chosen	536 (21)	-4	521 (46)	1.95 (0.04)
		Not Chosen	540 (26)		406 (48)	1.67 (0.04)
	Different Reliable	Chosen	511(23)	6	500(54)	1.94 (0.05)
		Not Chosen	505 (21)		347(39)	1.66 (0.04)
	Different Not reliable	Chosen	556(31)	56	462 (49)	1.90 (0.05)
		Not Chosen	500 (28)		416 (49)	1.75 (0.06)