Title
Beyond probability gain: Information access strategies in category learning

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 33(33)

ISSN
1069-7977

Authors
Meier, Kimberly
Blair, Mark

Publication Date
2011

Peer reviewed

Beyond probability gain: Information access strategies in category learning

Kimberly M. Meier (kim_meier@sfu.ca)
Mark R. Blair (mark_blair@sfu.ca)

Cognitive Science Program & Department of Psychology,
Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada V5A 1S6

Abstract

The present study uses eye-tracking to study information access in the context of category learning. Prior research has pointed toward the importance of *probability gain*, the increase in the chance of getting an answer correct, as a key variable in determining what information is considered most useful to acquire before making a classification decision. We manipulate the probability gain of three features in a four-category learning task by changing the base rates of the categories to be learned. Using participants' eye-movements to determine the order in which they acquire information after many trials of training, we find that increasing the probability gain of a feature does bias participants' first fixation. However, participants' strategies for acquiring feature information indicate they are more sensitive to efficiency goals: even with the low cost of eye-movements, participants direct attention to maximize efficiency, and do so without trading-off accuracy.

Keywords: attentional learning, categorization, eye-movements, information acquisition, probability gain.

Introduction

The category learning paradigm offers ideal conditions for studying how humans learn to differentially attend to information, because it encompasses the cognitive skills exercised by people in complex situations. Categorization tasks incorporate perception, attention, decision-making, memory, and motor behavior in a well-controlled environment within the one-third to three second timeframe argued to underlie the basic mechanisms that contribute to embodied cognition (Gray, Sims, Fu, & Schollelles, 2006). These tasks also measure changes in behavior across time, in response to experience. To solve a categorization task, participants must be able to figure out which features of a stimulus are useful for making classification decisions.

Recent work by Nelson, McKenzie, Cottrell, and Sejnowski (2010) suggests that participants evaluate the usefulness of a diagnostic feature by its *probability gain*. Probability gain defines the utility of looking at a feature as the extent to which checking this feature increases the probability, above chance, of making a correct decision; it assumes perfect knowledge of the total category probabilities and their features. Formally,

$$probabilityGain(F) = \left(\sum_{f_j} P(f_j) \times \max_{c_i} P(c_i \mid f_j)\right) - \max_{c_i} P(c_i)$$

where F represents a feature of unknown form to be looked at; f_j represents the specific forms a feature can take; and c_i represents specific categories (for example calculations, see Nelson, 2005). Note that maximizing probability gain corresponds to minimizing error; in fact, error-reduction has been proposed as the mechanism for driving shifts in attention during learning (Kruschke, 2003).

Using a probabilistic framework, under which beliefs about two categories are updated using Bayes's theorem, Nelson et al. (2010) investigated probability gain and three other plausible utility functions (information gain, Kullback-Liebler distance, and impact) for evaluating which of two features has the highest expected usefulness. Participants were given lengthy training in one of four tasks, each with different feature likelihoods, before completing a testing phase in which they could reveal only one of two probabilistic category features before making a classification decision. In all cases, and in a subsequent experiment, the majority of participants selected the feature that maximized probability gain.

While these experiments provide convincing evidence that probability gain is the utility function of choice in the kind of task Nelson et al. (2010) investigated, it is not clear yet how widely this finding can be generalized to more realistic tasks. In the paradigm used by Nelson et al., participants are only able to access a single piece of information. Typically, though, people have several available sources of information and they may wish to sample more than one. For example, in medical diagnostics scenarios it might be that a combination of less diagnostic tests has higher utility than a single, more diagnostic test. This raises the question of whether probability gain should be calculated for each feature sampled, or for each combination of features sampled. Another factor worth considering is information access costs: in real-world situations, information invariably has associated costs, whether they be time, money, or effort. Such factors may weigh against the feature with the highest probability gain; for example, one medical test may be slightly more effective, but vastly more expensive.

The goal of the present study is to investigate the extent to which probability gain plays a role in contexts where information access is not restricted by a single mouse click; instead, participants are allowed to sample multiple sources. To avoid large information access costs, as those have been shown to influence information access strategies (e.g., Gray & Fu, 2004; Wood, Fry & Blair, 2010), we chose to record participants' gaze with eye-trackers. Using eye-tracking to investigate how people attend to features when learning categories has gained recent popularity (e.g., Rehder & Hoffman, 2005; Kruschke, Kappenman, & Hetrick, 2005). This method has the advantage of providing a measure of participant behaviour *during* a categorization trial, allowing us to monitor how participants acquire information when making decisions. Though not free, eye-movements have some of the lowest time and energy costs possible.

In the present study we teach participants to identify four categories of fictitious micro-organisms with three binary-valued features (see Figure 1). All three features are relevant for successful performance in the task; however, only two features are relevant for making any one classification. As can be seen in the figure, feature 1 (F1) indicates whether the stimulus is a group A or group B category; feature 2 (F2) distinguishes between A1 and A2; and feature 3 (F3) distinguishes between B1 and B2. Unlike Nelson et al. (2010), features are fully diagnostic of category membership, rather than probabilistic.

We manipulated probability gain across two between-subjects conditions by altering the presentation frequency of A and B categories. In the 1:1 condition of this experiment, all four categories are presented with equal frequency (25%). In the 5:1 condition, we present five group A category stimuli for every group B stimulus. Thus, the probability of sampling a member from a particular A category is 41.67%, and 8.33% for each B category. The utility of sampling a feature is therefore different in the two conditions: it is equal (.25) for all three features in the 1:1 condition, but in the 5:1 condition the utility of F2 (.4167) is much higher than F1 (.0833) or F3 (0). To ensure participants had a good understanding of the categories and feature probabilities, we ran an experiment with 480 trials; typically, this kind of task is learned in 100 (Blair, Watson, & Meier, 2009).

The manipulation of probability gain should have a strong impact on how likely it is for participants to fixate a feature after learning the categorization task. If the decision to access a feature is based solely on probability gain, then certain predictable patterns will emerge. Participants in the 1:1 condition will show no preference for any dimension, as they all have equal probability gain. Participants in the 5:1 condition will fixate F2 before F1, as it has a much higher probability gain. However, if participants are sensitive to time, even under circumstances of very low information access costs, then we may see different patterns emerge. Fixating either F2 or F3 at the beginning of a trial runs the risk of spending time on information irrelevant for that trial. In both conditions, participants can minimize their number of fixations by fixating F1 first, then F2 if F1 is consistent with group A categories, or F3 if it is consistent with group B categories. Alternately, participants may be very motivated to speed things along, in which case we may see participants conserving fixations even though it negatively impacts accuracy in the task. This seems most likely in the 5:1 condition, where F2 alone is diagnostic of the correct category 83% of the time.

Method

Participants

134 students at Simon Fraser University participated in this experiment for course credit. All participants had normal or corrected-to-normal vision.

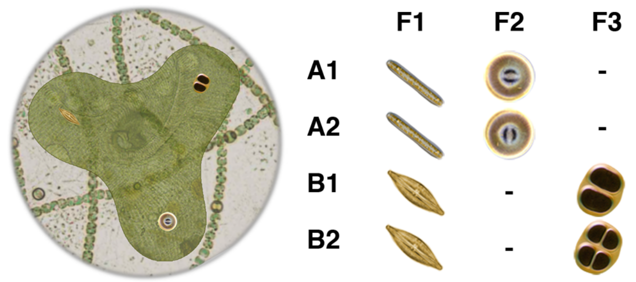

Figure 1: An example stimulus (left) and the category structure with example features (right). Features take one of two forms. Dashes indicate feature values are not useful for classification. Participants in the 1:1 condition saw an equal presentation of category members; participants in the 5:1 condition saw five group A categories for every group B.

Stimuli and Categories

Participants classified images of fictitious organisms. Each organism was composed of three features, one in each of the micro-organism's lobules (Figure 1, left). At a viewing distance of 70 cm. the area of each feature subtended 1.3º of visual angle, and features were equidistant from the centre of the organism, approximately 10.6º apart from one another. These features combined to form the category structure illustrated in Figure 1 (right). Feature images, locations and category labels were held constant for individuals and counterbalanced across participants.

Participants were assigned to one of two conditions. The 1:1 condition consisted of an equal presentation rate of A and B categories, and the 5:1 condition consisted of five common categories for every rare category. Note that the relative frequency of categories in the 5:1 condition was counterbalanced such that half of the participants saw common group A categories, and half saw common group B categories; for clarity, results are reported as though all participants viewed more group A categories.

Procedure

The procedure for the category learning task was identical in both the 1:1 and the 5:1 conditions, with the exception of presentation frequency noted above. Learning occurred by trial-and-error through corrective feedback. Both conditions included 480 trials. Each block of 24 trials was separated by a short break that indicated accuracy on the previous block as well as the number of blocks remaining.

Each trial began with a centrally-presented fixation cross. After pressing a button to advance the trial, the participant was shown an organism to classify. Participants had as much time as they liked to view the stimulus, and indicated responses on a four-button gamepad. Performance feedback was presented for 500 ms as a solid green (correct) or red (incorrect) mask, and the participant's response and correct category label were revealed in the centre of the screen as the organism reappeared. Participants inspected the labelled organism for as long as they wished before pressing a button to advance to the next trial.

Gaze Data Collection and Analysis

Gaze was recorded with a Tobii X120 eye-tracker sampling at 120 Hz with a spatial resolution of 0.5º. Participants with excessive sampling failures (defined as more than 30% of total samples lost) were discarded from analyses. Gaze data were transformed into fixations with a modified dispersion threshold algorithm (Salvucci & Goldberg, 2000) using spatial and temporal thresholds of 1.9º and 75 ms. A fixation was counted to a feature if it fell within 140 pixels of a feature's centre. To correct for posture changes over time, absolute values of fixation locations were corrected against fixations to the central cross at the beginning of each trial. Gaze analyses were conducted only on trials for which less than 25% of gaze points recorded during that trial were lost.

Results

Of the 134 students who participated in the experiment, six failed to complete the task. Data from an additional 12 participants were excluded due to an excessive number of eye-tracker sampling errors. Four participants were discarded for responding randomly throughout the task.

Because we are interested in the behaviour of subjects who have complete knowledge of the category structure by the end of the experiment, analyses are conducted only on the 43 of 54 participants in the 1:1 condition and 52 of 58 participants in the 5:1 condition who reached a learning criterion of 24 correct responses in a row. The mean number of trials to reach this criterion was 133.77 (SD = 96.63) for the 1:1 condition, and 117.00 (SD = 99.03) for the 5:1 condition ($t < 1$). Many of the current analyses focus on trials during the last quarter, or 120 trials, of the experiment. For participants in the 1:1 and 5:1 conditions, respectively, accuracies during these trials were .98 (SD = .02) and .96 (SD = .04).

Of the six participants in the 5:1 condition who did not reach the learning criterion, two appeared to learn common categories (with accuracies of .92 and .81 over the last quarter) while completely ignoring rare ones (accuracies of 0). The remaining four appeared to be slow learners. Although we do not consider them here, the two who learned common but not rare categories usually fixated F2-only, and had fast response times. The surprising fact that only two participants were willing to trade accuracy for decreased time and effort costs indicates that participants were highly motivated to do well in this task.

First Fixations

If participants are using probability gain as the basis for viewing stimulus features, we would expect this to show up most strongly in the first fixation. To understand how participants learned to access information for solving the task over time, we plotted the mean proportion of participants' first fixations to F1, F2, and F3 in each condition (Figure 2). Participants in the 1:1 condition, where probability gain is .25 for all features, clearly learn to fixate F1 first: over the last quarter of the experiment, 74% of all trials begin with a fixation to this feature, while 12% and 13% of all trials begin with fixations to F2 and F3. As noted in the introduction, gathering information from F1 first allows participants to make the fewest number of fixations while maintaining perfect accuracy. Participants in the 5:1 condition also tend to fixate F1 first, but the high probability gain of F2 in this condition (.4167, vs .0833 for F1) appears to be boosting first fixations to this location: over the last quarter, 57% of trials begin with a fixation to F1, and 41% begin with a fixation to F2. Only 2% of all trials begin with a fixation to F3. Participants' mean proportion of F1-first trials is lower in the 5:1 condition (M = .37, SD = .25) than in the 1:1 condition (M = .50, SD = .27), $t(93) = 2.28$, p = 0.02. The mean proportion of F2-first trials, alternately, is higher in the 5:1 condition (M = .26, SD = .24) than in the 1:1 condition (M = .081, SD = .15), $t(93) = 4.32$, $p < .001$.

In general, F1-first and F2-first trials over the last quarter of the experiment in the 5:1 condition differ only slightly. On average, fewer fixations are made on trials that begin with fixations to F2 (M = 2.40, SD = 1.75) than to F1 (M = 2.53, SD = 1.61), $t(3966) = 2.32$, p = .02; although response time on F1-first trials (M = 1363 ms, SD = 793) is no faster

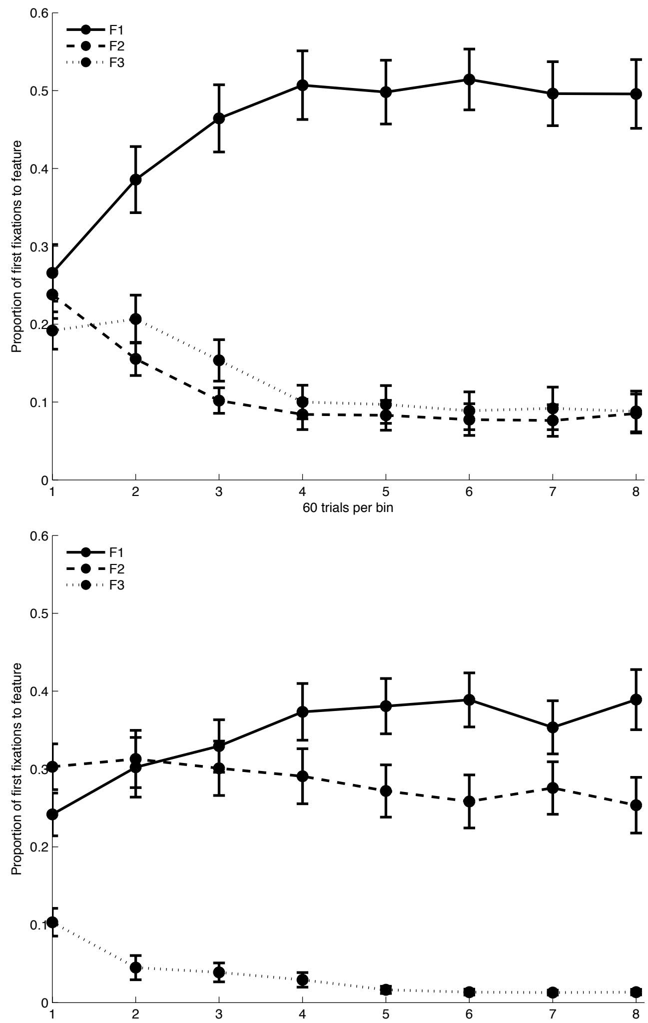

Figure 2. Mean proportion of fixating each feature first in eight bins across the experiment for participants in the 1:1 condition (top) and 5:1 condition (bottom).

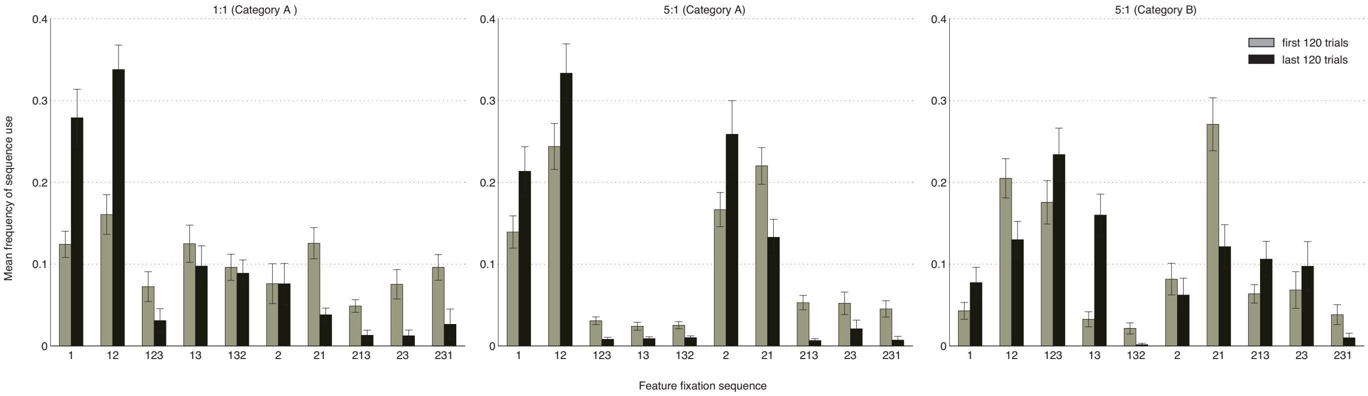

Figure 3. Frequency histograms displaying the mean frequency of each strategy's use by participants in the 1:1 condition, category A (left); and in the 5:1 condition, common category A (middle) and rare category B (right); for the first and last 120 trials of the experiment. Error bars reflect standard error of the mean. Note that over a block of 120 trials, the 1:1 condition contains 60 category A presentations; the 5:1 condition contains 100 category A, and 20 category B, presentations. F1 distinguishes between A and B categories; F2 is diagnostic of group A categories, and F3 is diagnostic of group B categories.

than F2-first trials (M = 1376 ms, SD = 862; t < 1). Accuracy on both trial types is high, though significantly different, $t(3966) = 4.68$, $p < .001$; for F1-first trials, M = 0.98 (SD = .15) and for F2-first trials, M = .95 (SD = .22).

Information-access sequences

To better understand how participants' strategies for acquiring information from the three available features are affected by presentation frequency, we investigated the relative frequency of ten possible information sequences; that is, ten orders in which participants can fixate category features. Given the low number of trials that begin with fixations to F3 in the 5:1 condition (recall this feature is diagnostic of rare group B categories only), we restrict our analysis to sequences that begin with F1 or F2. We also collapse repeated fixations to features, as we are primarily interested in the order in which information is gathered.

Figure 3 displays the mean relative frequency of each information sequence occurring, for the first and last quarter of the experiment, in group A categories for participants in both conditions and group B categories in the 5:1 condition. Recall that accurately classifying group A categories requires F1 and F2; while group B requires F1 and F3. (Because we are not looking at patterns beginning with F3, we are omitting group B categories in the 1:1 condition; F3 data for these categories are complementary to the F2 data for group A categories.) Table 1 summarizes mean frequency, accuracy, and response time for each strategy in the last quarter of the experiment. Correlations reported below use participants' mean accuracy and response time over these last 120 trials.

Fixation patterns - 1:1 condition. While viewing group A categories, participants in the 1:1 condition (Figure 3, left) tend to use a diverse set of strategies in the first quarter of the experiment. By the last quarter of the experiment, the most common sequence is fixating F1 first, and F2 second. Use of this sequence is correlated with F1-F3 sequences during group B categories, $r(41) = .412$, $p = .006$; indicating that participants are accessing F1 first in order to sample the most appropriate feature second. This strategy, in theory, allows participants perfect accuracy while making the fewest fixations possible. However, increased use of F1-F2 sequences has no relationship with accuracy or response time (ps > .05).

Surprisingly, participants commonly fixate F1 only. This pattern should not yield enough information for accurate performance. Closer inspection reveals that participants appear to be covertly attending to features they do not fixate: that is, they are able to direct attention towards a feature without initiating an eye-movement. The result is a very fast, very accurate trial. F1-only trials are, on average, about 450 ms faster than F1-F2 trials, and accuracy on these trials appears no worse (Table 1); indeed, participants who use a higher proportion of F1-only sequences are faster at

Table 1. Mean frequency (Freq), accuracy (Acc), and response time in ms (RT) for trials of each sequence during the last 120 trials of group A categories in the 1:1 condition, and A (common) and B (rare) categories in the 5:1 condition.

	1:1 (A)			5:1 (A)			5:1 (B)		
Order	Freq	Acc	RT	Freq	Acc	RT	Freq	Acc	RT
1	.28	.98	1259	.21	.98	1051	.08	.98	1330
12	.34	.99	1702	.33	.97	1623	.13	.96	1639
123	.03	1	3095	.01	.98	2391	.23	.93	1738
13	.10	.92	2024	.01	.99	1599	.16	.96	1548
132	.09	.96	1872	.01	1	1756	0	.96	1607
2	.08	1	1471	.26	.95	1073	.06	.69	1276
21	.04	1	1978	.13	.98	1486	.12	.76	2044
213	.01	.83	2156	.01	.96	1897	.11	.91	1842
23	.01	.98	3227	.02	1	2303	.10	.83	2067
231	.03	1	3841	.01	1	3060	.01	1	5451

responding: $r = -.361$, $p = .017$; but use of this strategy has no negative relationship with accuracy, $p > .1$. Participants who tend to use the F1-only strategy when viewing group A categories also use this strategy when viewing group B categories, $r = .705$, $p < .001$.

Other participants also seem to be using highly efficient strategies employing covert attention. As seen in Table 1, F2-only is not as common as F1-only, but shows the same pattern of quick (about 230 ms faster than F1-F2 trials) and accurate responses. Participants who rely more on this strategy tend to use the covert F3-only strategy when viewing group B categories, $r = .736$, $p < .001$.

In all, the mean proportion of single-feature fixation trials increases from .18 ($SD = .13$) in the first quarter to .32 ($SD = .20$) in the last quarter of the experiment, $t(42) = 4.71$, $p < .001$. A higher proportion of single-feature trials over this last quarter is correlated with faster response times, $r = -3.97$, $p = .008$; but has no relationship with accuracy, $p > 0.3$. In other words, participants who adopt covert strategies are saving time without sacrificing accuracy.

Though uncommon, F1-F3 and F1-F3-F2 sequences also occur, which likely reflect the kinds of fixation errors people make when trying to distribute attention dynamically in response to information gathered during a first fixation.

Fixation patterns - 5:1 condition. Unlike participants in the 1:1 condition, participants in the 5:1 condition appear to use fewer distinct strategies for accessing information during the first quarter of the experiment when viewing their common group A categories (Figure 3, middle). Both F1-F2 and F2-F1 sequences are often used, as are F1-only and F2-only. While participants continue to rely on three of these strategies through the last quarter of the experiment, the mean frequency of F2-F1 significantly decreases, $t(51) = 3.83$, $p < .001$. The sequences used during common categories in the first quarter of the experiment are also found while viewing rare group B categories (Figure 3, right), with the clear addition of the sequence F1-F2-F3. In the last quarter of the experiment, participants increase their reliance on the F1-F3 sequence, $t(51) = 5.03$, $p < .001$; and decrease their reliance on the F2-F1 sequence, $t(51) = 4.24$, $p < .001$.

There appear to be two general overt strategies for acquiring feature information: F1-F2, or F2-F1; followed by a fixation to F3 if the stimulus is a rare one. Greater reliance on F1-F2 during common categories is most strongly correlated with use of F1-F2-F3 sequences during rare categories, $r(50) = .758$, $p < .001$; and is also associated with F1-F2 ($r = .304$, $p = .028$) and F1-F3 ($r = .340$, $p = 0.014$) sequences. On the other hand, participants more often relying on F2-F1 during common categories also rely on this sequence during rare categories ($r = .689$, $p < .001$); and on F2-F1-F3 sequences ($r = .502$, $p < .001$). The degree to which participants rely on F1-F2 or F2-F1 during common categories has no association with response time (ps $> .1$) but greater use of F1-F2 sequences is correlated with higher accuracy ($r = .321$, $p = .02$).

Single-feature fixation strategies, that is, strategies that rely on covert attention, also appear in the 5:1 condition. Although the mean response times associated with F1-only and F2-only trials are similar, and about 400 ms faster than the next quickest response time (Table 1), only participants with a higher proportion of F2-only trials gain a response time advantage, $r = -.375$, $p = .006$. Neither strategy appears to help or hurt mean accuracy (ps $> .1$).

As in the 1:1 condition, participants appear to increase their use of covert strategies in order to save time without impeding their task accuracy. The mean proportion of single-feature fixation trials increases from .25 ($SD = .02$) in the first quarter of the experiment to .40 ($SD = .03$) in the last quarter of the experiment, $t(51) = 4.76$, p $< .001$. Increased use of single-fixation trials has no relationship with accuracy ($p > .5$), but is correlated with faster response times, $r = -.375$, $p = .006$.

Discussion

The present study investigated the impact of probability gain on fixation sequences in a categorization task. One key finding is that the usefulness of information, as measured by probability gain, has a significant and lasting influence on strategies for accessing this information. The probability gain of F2 in the 5:1 condition (.4167) is higher than the probability gain of F2 in the 1:1 condition (.25), and indeed, a higher proportion of fixations were made to F2 first in the 5:1 than in the 1:1 condition. However, it is also clear from our data that probability gain was not the only factor which influenced information access, and there were several clear indicators that participants were deploying attention more efficiently than would be necessary to maximize probability gain alone.

One indicator was that fixating the highest probability gain feature first was not the dominant strategy of participants in either condition. In the 1:1 condition, where the probability gain of sampling any of the three feature was equal (.25), participants overwhelmingly chose to fixate F1 first; thus, efficiency considerations were enough to push fixations toward a particular feature. By fixating F1 first, participants were able to minimize the number of eye-movements required to gather the necessary information for perfect accuracy. In the 5:1 condition the situation is more dramatic: even where the probability gain of F1 (.0833) is far lower than F2 (.4167), there were still more first-fixations to F1 than to F2 for most of the experiment. This finding corresponds well with work by Matsuka and Corter (2008). In a medical diagnostic task where participants can reveal only one feature at a time with mouse-clicks, they found, under a number of conditions, that participants used cost-effective strategies for accessing feature information. We extend this idea by showing that not only *which* features are selected, but also the *order* in which features are selected, are both important considerations for participants, even at the level of eye-movements.

The high prevalence of single-feature fixation trials, which increased over time, also emphasizes the importance of efficiency. This may be in part a function of expertise, as participants have a long time to practice making more efficient fixations. Reanalysis of previously published data in which participants performed a task similar to the 1:1 condition but for only 96 trials after reaching a learning criterion (Blair, Watson, & Meier, 2009) indicates a low mean proportion of F1-only strategies (.029) during these trials. In the current study, we find participants increasing their use of single-feature fixation strategies over a 480-trial experiment. Participants using these strategies have, in general, gained a response time advantage without trading-off accuracy. There is evidence that covert attentional shifts to a target location precede voluntary saccades to the location (Kowler, Anderson, Dosher, & Blaser, 1995; McPeek, Maljkovic, & Nakayama, 1999), so it may be that practiced participants are sensitive enough to visual information at feature locations that they can respond before initiating an eye-movement to these locations.

In one sense, the general finding that people prefer the most efficient method of achieving a goal is not surprising. Even in the context of low-cost eye-movements, it has already been shown, for example, that adult readers have more efficient fixation patterns than children (Rayner, 1985). Our task, however, is not a task at which participants have trained for many years. Here, participants are learning a novel task, and given about an hour of practice. The task is self-paced, with no emphasis on speed. Despite this, participants come up with ways to improve their efficiency, while still answering accurately.

Finally, our findings lend support to the idea that eye-movements can be dynamically, and consistently, deployed to information in real-time. In the 1:1 condition, participants would often sample F1, and use information at this location to direct eye-gaze to F2 or F3. This strategy, and a variety of others (eg., F2-F1 and F3 if necessary; F1-F2 and F3 if necessary), also emerged in the 5:1 condition. Participants adopting F1-F2 / F1-F3 strategies are making fewer fixations, but these strategies do not appear to save time. There was surprisingly little difference in response times between most of the overt information-access strategies. This suggests that in this context, repeated practice of a sequence – be it one beginning with F1, or one beginning with F2 – helps efficiency as much as dynamically changing fixation order. Perhaps the time it takes to decide where to fixate next is roughly equivalent to the time spent occasionally fixating an unnecessary feature in a well-practiced fixation sequence.

Nelson et al. (2010) have shown that "information acquisition optimizes probability gain." Our data confirm the importance of probability gain, but also suggest a qualifier: *all things being equal.* Our results show that when given the opportunity to sample any or even every available source of information, participants use more than just immediate probability gain for deciding what feature is most important to sample. Instead, the deployment of overt attention reflects a more complex strategy which is also sensitive to the cost of accessing information. Participants with minimal training can retain accuracy while being remarkably frugal in how they acquire information. This seems to be true even when, as with eye-movements, information is cheap. The present work suggests that choosing which information to access is a function of both its utility and its cost.

Acknowledgments

This research was supported by grants to MRB from the National Science and Engineering Research Council of Canada, the Canada Foundation for Innovation, and the British Columbia Knowledge Development Fund. We would like thank the Cognitive Science Lab team for all their help.

References

Blair, M.R., Watson, M. R., & Meier, K.M. (2009). Errors, efficiency, and the interplay between attention and category learning. *Cognition, 112*, 330-336.

Gray, W. D., & Fu, W.-T. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science, 28*, 359-382.

Gray, W. D., Sims, C. R., Fu, W.-T., & Schollelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review, 113*, 461-482.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research, 35*, 1897-1916.

Kruschke, J. K. (2003). Attention in learning. *Current Directions in Psychological Science, 12*, 171-175.

Kruschke, J. K., Kappenman, E. S., & Hetrick, W. P. (2005). Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. *Journal of Experimental Psychology: Learning, Memory & Cognition, 31*, 830-845.

Matsuka, T. & Corter, J. E. (2008). Observed attention allocation processes in category learning. *Quarterly Journal of Experimental Psychology, 61*, 1067-1097.

McPeek, R.M., Maljkovic, V., & Nakayama, K. (1999). Saccades require focal attention and are facilitated by a short-term memory system. *Vision Research, 39*, 1555-1566.

Nelson, J. D. (2005). Finding useful questions: On bayesian diagnosticity, probability, impact, and information gain. *Psychological Review, 112,* 979-999.

Nelson, J. D., McKenzie, C. R. M., Cottrell, G. W., & Sejnowski, T. J. (2010). Experience matters: Information acquisition optimizes probability gain. *Psychological Science, 21,* 960-969.

Rayner, K. (1985). The role of eye movements in learning to read and reading disability. *Remedial and Special Education, 6*, 6-53.

Rehder, B. & Hoffman, A.B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology, 51*, 1-41.

Salvucci, D. D. & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the Eye Tracking Research and Applications Symposium* (pp. 71-78). New York: ACM Press.

Wood, M. J., Fry., M. D., & Blair, M. R. (2010). The price is right: A high information access cost facilitates category learning. *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (pp. 236-241). Austin, TX: Cognitive Science Society.