Predicting Learned Inattention from Attentional Selectivity and Optimization

Abstract

Although selective attention is useful in many situations, it also has costs. In addition to ignoring information that may become useful later, it can have long term costs, such as learned inattention – difficulty in learning from formerly irrelevant sources of information in novel situations. In the current study we tracked participants' gaze while they completed a category learning task designed to elicit learned inattention. During learning an unannounced shift occurred such that information that was most relevant became irrelevant, whereas formerly irrelevant information became relevant. We assessed looking patterns during initial learning to understand how different aspects of attention allocation contribute to learned inattention. Our results indicate that learned inattention depends on both the overall level of selectivity (measured as entropy of proportion of looking to each feature) and the extent to which participants optimized attention (becoming more selective over time).

Keywords: selective attention; categorization; learning; attention

Introduction

Category learning is a critical cognitive process that enables abstract thought and allows for generalization of knowledge to novel situations. Since the early work by Shepard, Hovland, and Jenkings (1961) selective attention has been considered a critical component of categorization and category learning. Selective attention refers to the ability to prioritize task-relevant information and filter out task irrelevant information (Desimone & Duncan, 1995; Hanania & Smith, 2010; Plude, Enns, & Brodeur, 1994; Pashler, Johnston, & Ruthruff, 2001; Yantis, 2000).

Most models of categorization and category learning adopt the Shepard et al. (1961) view and consider selective attention a critical contributor to human categorization. Exemplar models (Hampton, 1995; Medin & Schaffer, 1978; Nosofsky, 1986), prototype models (Smith & Minda, 1998), clustering models (Love, Medin, & Gureckis, 2004), and dual process models (Ashby, Alfonso-Reese, Turken, & Waldron, 1998) all include some form of selective attention as a factor determining the influence (or weight) of stimulus dimensions on categorization. According to some of these models, as participants learn categories, they tend to shift attention to features that are more likely to predict category membership, while attending less to less predictive features, the idea known as attention optimization. This idea has been confirmed empirically: there is eye-tracking evidence indicating that when learning categories, people indeed tend to optimize their attention over time, increasing fixating at the features most predictive of a category and decreasing fixating at irrelevant features (Rehder & Hoffman, 2005; Blair, Watson & Meier, 2009).

Costs and Benefits of Selectivity

Although selective attention is often beneficial in many learning scenarios (e.g. faster, efficient processing of attended information), selective attention also has costs (Best, Yim, & Sloutsky, 2013; Hoffman & Rehder, 2010; Plebanek & Sloutsky, 2017; Rich & Gureckis, 2018). Some of the costs are relatively short-term: people miss non-selected information. Other costs are longer term in that they affect future learning. One such short-term cost is that non-selected information is filtered out. Focusing attention is a tradeoff in that it results in efficient learning and performance, but it also results in missing information that could be used later.

While short-term costs of selectivity affect only the task at hand, longer term costs also affect performance on future tasks. One type of long-term cost has been recently discussed by Rich & Gureckis (2018), who referred to it as a "learning trap". The authors demonstrate that under certain circumstances selective attention can be a trap in that it can result in getting stuck on inaccurate representations of the to-be-learned structure and preventing the exploration needed to discover the correct structure. This happens particularly when there are possible negative outcomes to exploring, in which case selective attention results in overgeneralizing which things should be avoided.

A more general long-term cost is that selective attention may result in learned inattention (see Hoffman & Rehder, 2010, for a review) – difficulty in learning from sources of information that were uninformative in a previous situation. Optimizing one's attention to the currently most relevant sources of information can result in not only learning to ignore currently irrelevant sources of information, but continuing to ignore these sources in novel situations in the future (Kruschke & Blair, 2000). As a result, if those sources of information become relevant later, learning is more difficult than it would have been if one had not first learned to ignore them. Learned inattention can be detrimental when task demands change, or when a new classification contrast depends on previously irrelevant features.

For example, Hoffman & Rehder (2010) had participants learn to classify stimuli consisting of three dimensions. Learning occurred in either a classification condition or a feature inference condition. In classification, participants predict the category label from all of the stimulus' features. In inference they predict one missing feature from the label and the remaining features.

In the first phase of the experiment only dimension 1 was relevant to distinguish two categories, while the other two were irrelevant. In a second phase, only dimension 2 was relevant for distinguishing two new categories, while in a third phase dimension 3 was relevant for a novel contrast between categories. Classification encourages selective attention since only a single feature predicts category membership. Inference, on the other, encourages distributed attention, since participants may need to predict any of the three features.

Participants performing the classification task were impaired at learning the new contrast when the relevant dimension changed compared to participants doing the feature inference task. Additionally, eye-tracking showed that the classification participants were much less likely to fixate the relevant dimension after learning it was irrelevant in a prior phase of the experiment compared to baseline levels at the start of the experiment. These costs occurred because learners selectively attended the relevant feature while classifying the stimuli—optimizing their attention to ignore (or inhibit) the other features.

The Current Study

The goal of this study is to further investigate learned inattention during category learning by examining how different aspects of attention allocation contribute to learned inattention. This study also serves as a first step toward a larger investigation of developmental differences in attention allocation and optimization and their consequences. The current study investigates only adults, but developmental implications and predictions for children are touched on in the Discussion.

In our experiment, we presented participants with a category learning task while tracking their gaze. The to-be-learned categories had a rule-plus-similarity structure, with one deterministic feature perfectly predicting category membership and multiple probabilistic features, providing good, but imperfect prediction (see Deng & Sloutsky, 2016, for a similar structure). In addition, one feature was completely irrelevant to categorization.

Given the structure, participants could either form a rule-based representation (by relying on the deterministic feature) or a similarity-based representation (by relying on all features). So, either selective or distributed attention could lead to effective learning. Once participants mastered the categories in this initial phase, there was an unannounced shift in the category learning task. After the shift, feature dimensions that had been deterministically predictive in became irrelevant, and features that had been irrelevant in became deterministically predictive.

Learned inattention was expected to produce costs on learning in Phase 2, making participants less likely to attend to and learn to use the new deterministic feature. We examined what aspects of attention during initial learning were most likely to manifest these costs. In particular we investigated the effects of overall selectivity and of attention optimization (increase in selectivity over time).

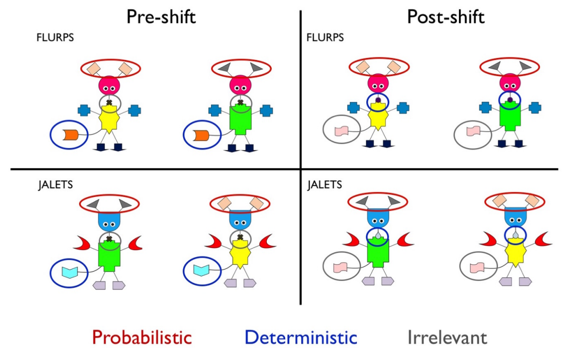

Figure 1. Examples of stimuli. The stimuli were creatures composed of seven discrete-valued dimensions differing in shape and color. One Deterministic feature perfectly predicted category membership (the tail in this example). One Irrelevant feature was the same across both categories (the button on the neck here). Five Probabilistic features predicted category membership with 80% accuracy. After an unexpected shift occurred, the Deterministic became Irrelevant, and the Irrelevant feature became Deterministic Probabilistic features were unchanged by the shift.

Method

Participants

A total of 38 adults (26 women) participated in the experiment. Participants were undergraduate students participating for course credit.

Materials and Design

Stimuli were colorful images of creatures composed of seven discrete-valued dimensions (see Figure 1). The creatures were divided into two categories, referred to as Flurps and Jalets. Of the seven features (antenna, head, body, button, hands, feet, and tail), one feature deterministically predicted category membership (henceforth the Deterministic feature), five features were probabilistically predictive with 80% accuracy (the Probabilistic features), and one feature was non-predictive—having the same value across all exemplars of both categories and therefore irrelevant to classification (i.e., the Irrelevant feature). Table 1 shows the stimulus structure used in the task.

Stimuli were organized into pairs of complementary sets. Each set in a pair was identical to its counterpart except that the Deterministic and Irrelevant features swapped roles. Probabilistic feature values and the category labels remained the same. As discussed below, participants learned one set in Phase 1 of the experiment (i.e., before the shift), and then the stimuli were unexpectedly replaced with the complimentary set for Phase 2 (i.e., after the shift). There were two pairs: one where feet and hands were the Deterministic/Irrelevant features, one where tail and neck button were the Deterministic/Irrelevant features. Which pair was presented and which set in that pair was learned in Phase 1 were counterbalanced between participants.

Stimulus sets also contained Ambiguous items. These were hybrid items having the Probabilistic features from one category and the Deterministic feature from the other category. These items were presented only during the testing sessions and were designed to test which features controlled categorization. There were 10 Ambiguous items total per set—one corresponding to each exemplar seen during training (i.e. with identical Probabilistic features but with the Deterministic feature from the opposite category). These items allow us to determine whether participants' category judgments were based more on the single Deterministic feature or on one or more of the Probabilistic features.

Table 1. Stimuli structure during training

	Pre-shift							Post-shift							
Feature	D/I	I/D	P1	P2	P3	P4	P5		D/I	I/D	P1	P2	P3	P4	P5
Category A															
A1	1	2	0	1	1	1	1	A6	2	1	0	1	1	1	1
A2	1	2	1	0	1	1	1	A7	2	1	1	0	1	1	1
A3	1	2	1	1	0	1	1	A8	2	1	1	1	0	1	1
A4	1	2	1	1	1	0	1	A9	2	1	1	1	1	0	1
A5	1	2	1	1	1	1	0	A10	2	1	1	1	1	1	0
Category B															
B1	0	2	1	0	0	0	0	B6	2	0	1	0	0	0	0
B2	0	2	0	1	0	0	0	B7	2	0	0	1	0	0	0
B3	0	2	0	0	1	0	0	B8	2	0	0	0	1	0	0
B4	0	2	0	0	0	1	0	B9	2	0	0	0	0	1	0
B5	0	2	0	0	0	0	1	B10	2	0	0	0	0	0	1

Note: D/I is the feature that is Deterministic prior to the shift and Irrelevant after the shift. I/D is the feature that is Irrelevant pre-shift and Deterministic post-shift. P1-P5 are the probabilistic features.

Procedure

Adult participants conducted a classification procedure while their gaze was monitored with an EyeLink 1000 hydraulic-arm eyetracker at 500Hz (SR research, Ontario, Canada). The experiment was divided into two phases. Both phases contained a training section (with feedback) followed by a testing section (no feedback). In Phase 1 participants learned to classify two categories of creatures (and then were tested), and then in Phase 2 an unannounced shift occurred wherein the previously Deterministic feature and the previously Irrelevant feature swapped roles.

The formerly Deterministic dimension took on a new, previously unseen, value that was fixed across all stimuli of both categories, while the formerly Irrelevant dimension now had two new potential values that perfectly predicted category membership. Participants were given no warning that this shift would occur at any point. Like Phase 1, Phase 2 consisted of training followed by testing.

At the beginning of the experiment participants were given information about the Deterministic and Probabilistic features. For Probabilistic features they were told that most of the members of the category had that particular feature value. For the Deterministic feature they were told that all members of category A have one value while those of category B have another value, while being shown both. The Irrelevant feature was never mentioned in the instructions. These informative instructions were included to ensure good learning in Phase 1, since any expected costs due to the shift rely on the categories initially being learned well. Additionally, as noted above we plan to eventually investigate developmental differences, and young children require this type of informative instructions and feedback in order to learn well within the timeframe of the experiment.

Training Training in each phase consisted of 30 trials (in 3 blocks of 10 trials). In each block of 10 trials the ten training exemplars, five from each category, were presented in random order, so participants saw each exemplar three times throughout training (see Table 1 and Figure 1).

On each training trial, one stimulus was presented in the middle of the screen and participants indicated which category they thought it belonged to. Corrective feedback was then given which tried to equally encourage attention to general appearance (similarity-based responding) and to the Deterministic feature (rule-based responding). For example feedback would be "Correct this is a Flurp. It looks like a Flurp and has the Flurp hands.", or "Oops this is actually a Jalet. It looks like a Jalet and has the Jalet hands.", in the case where hands were the Deterministic feature.

In Phase 2, after the unannounced shift, feedback was simplified to mention only the correct category without drawing attention to the features (e.g. "Correct this is a Flurp."). While this change in feedback may have given participants some indication that a change had occurred, it was necessary so that participants would need to figure out on their own the new contingencies between features and categories. That learning process in Phase 2, discovering what is informative, is the critical area of interest, while parity between Phase 1 and 2 is not critical.

Testing Testing in each phase consisted of 20 trials. Again, participants saw the stimuli one at a time and classified each one, but no feedback was provided during the test. The 10 items seen during training (henceforth the High Match items) and 10 Ambiguous items were each presented once, in random order. These two types of items, respectively, were the basis for the behavioral measures of general shift costs and shift costs due to learned inattention.

Accuracy on High Match items indicates how well each participant learned during the immediately prior training session. A decrease in accuracy from Phase 1 to Phase 2 on these items would indicate a general cost (in terms of poorer learning) due to the unexpected shift, which may occur for a number of reasons.

Responses to the Ambiguous items, in contrast, provide the cornerstone of our behavioral analyses related to learned inattention. Prior to the shift they provide the baseline level that each participant tended to categorize based on the single Deterministic feature. After the shift responses to the Ambiguous items tell us whether participants learned and used the rule on the new Deterministic (formerly Irrelevant) feature. Low deterministic responses indicate learned inattention since it suggests that the participant had difficulty finding the new rule on the feature that was previously irrelevant.

Results

Behavioral Results

Initial learning in Phase 1 was good overall (see Figure 2). Mean accuracy was 92.1% correct. A repeated measures ANOVA found a main effect of block [$F(1,75) =. 14.74$, $p = 0.0003$, partial $\eta^2 = 0.164$] suggesting that participants learned well during training. Categorization accuracy on High Match items during the test was also high ($M =$ 93.9% correct).

Responses to the Ambiguous items provide insight into which features controlled participants' categorization: Responding based on the Deterministic feature points to rule-based categorization, whereas responding based on the Probabilistic features suggests similarity-based (or at least non-rule-based) categorization. Distributing attention during training could lead to either type of responding (since all features were attended), but selective attention to the Deterministic feature should only result in rule-based categorization. Participants were overwhelmingly deterministic in their categorization of the Ambiguous items: 88.4% deterministic responses, which was well above chance, $t(37) = 9.69$, $p < 0.001$, $d = 1.57$.

Post-shift Learning Participants learned well in Phase 2 after the shift (77.0% correct), which was above chance, $t(37) = 19.09$, $p < 0.001$, $d = 3.10$, but accuracy was lower than prior to the shift, $t(37) = 6.91$, $p < 0.001$, $d = 1.12$, which is expected—even in the absence of learned inattention—due to less informative feedback in Phase 2 compared to Phase 1 and any general costs of adapting to the shift. A repeated measures ANOVA on Phase 2 training accuracy found a main effect of block [$F(1,75) = 7.192$, $p = 0.009$, partial $\eta^2 = 0.087$], suggesting accuracy increased over time. Accuracy on High Match items during the test was significantly higher than chance (M = 81.3% correct), $t(37) = 10.06$, $p < 0.001$, $d = 1.63$, but was also lower than in Phase 1, $t(37) = 3.52$, $p = 0.001$, $d = 0.571$, suggesting a learning cost due to the unexpected shift. This represents a general shift cost which may be due to a variety of factors including learned inattention. We assess costs specific to learned inattention in the next section.

Shift Costs Due to Learned Inattention We assessed effects of learned inattention by examining responses to the Ambiguous items in the Phase 2 test (Figure 3). Learned inattention would result in relatively low levels of deterministic responses on the post-shift test, since participants would be less likely to attend to the previously Irrelevant (now Deterministic) feature, and thus would have difficulty learning the new rule on that feature. This would result in participants relying primarily on the Probabilistic features instead when making category judgments after the shift. Participants were significantly below chance, M =36.6%, $t(37) = 2.53$, $p = 0.016$, $d = 0.41$, suggesting that they primarily relied on Probabilistic features to categorize after the shift, in contrast to their behavior prior to the shift.

Figure 3 shows individual participants proportion of classifying Ambiguous items based on the Deterministic feature in the post-shift test.

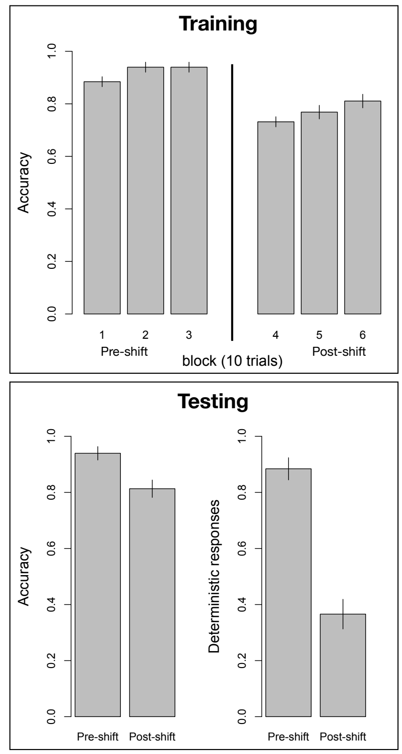

Figure 2. Behavioral results. During initial training participants learned well and achieved high accuracy. A substantial drop occurred after the shift. Accuracy during test was high for old (High Match) items, and was lower after the shift, displaying a general cost of the shift. Responding based on the Deterministic feature was very high prior to the shift, but dropped substantially in the post-shift test, suggesting effects of learned inattention. Error bars represent standard error of the mean.

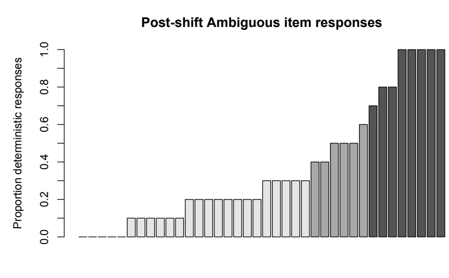

Figure 3. Individual participants' post-shift deterministic responding. Responses to Ambiguous items on the post-shift test varied widely between individuals. The majority classified based on the Probabilistic features, with some intermediate, and less than one-quarter of participants classifying based on the new Deterministic feature.

Eye-Tracking

Regions of interest isolating each feature were used to calculate the proportion of each trial spent looking at each feature for each subject. Timepoints where participants were not looking at any of the features were removed. Figure 4 shows the average proportion of looking at each feature type during training. Prior to the shift the proportion of the trial spent looking at the Deterministic feature increased over trials, while time spent looking at Probabilistic features decreased. Looking at the Irrelevant feature was extremely low throughout all of training (M = 2.82% of total looking time), but did decrease over the course of training—as measured by comparing block 1 (M = 3.60%) to block 3 (M = 1.52%), $t(37) = 3.39$, $p = 0.002$, $d = 0.161$.

After the shift, looking at the previously Deterministic (now Irrelevant) feature dropped off rapidly, while looking to the Probabilistic features shot up. Critically, looking at the newly Deterministic feature (that was previously Irrelevant) did not increase from block 1 to block 3, $t(37) = 0.53$, $p = 0.600$, demonstrating the effects of learned inattention. They continued to ignore this feature despite the high level of information it now contained.

We assessed attentional patterns during initial learning by calculating the entropy (Shannon, 1948) for each trial. Entropy was defined as,

$$S = -\sum_{i=1}^{n} p_i * \log(p_i)$$

where p_i is the proportion of the trial spent looking at feature i. Higher entropy indicates more distributed attention, where maximum entropy is produced by looking at all seven features equally, and lower entropy indicates more selective attention—focusing on a smaller number of features. We use entropy as a measure of selectivity rather proportion of looking at the Deterministic feature since some participants may have optimized their attention to one of the Probabilistic features, and that selectivity should still produce learned inattention despite being suboptimal. Entropy for each trial was normalized by dividing by maximum possible entropy, such that all values were between 0 and 1.

Drop in entropy was calculated as average entropy per trial in block 1 minus average entropy in block 3. This served as a measure of attention optimization, since greater drops indicate an increase in selectivity of attention.

We performed a logistic regression predicting classification of Ambiguous items on the post-shift test from average entropy during pre-shift training, the drop in entropy over training, and their interaction. This analysis revealed a significant interaction, $z = 3.318$, $p = 0.001$. To better understand the interaction, we divided participants into low and high entropy groups based on a median split. We then performed a logistic regression on Ambiguous item responses predicted from the drop in entropy for each group. For the low entropy group, there was a significant negative relationship between entropy drop and Ambiguous item responses, $z = -3.153$, $p = 0.002$, indicating that those who optimized their attention in the initial training were less likely to use the new Deterministic feature to categorize items after the shift (see Figure 5). In contrast, participants in the high entropy group did not show a relationship between the drop in entropy and responses to Ambiguous items, $z = 1.899$, $p = 0.0575$.

In other words, attention optimization was associated with greater learned inattention in the low entropy (selective attention) group, but not in the high entropy (distributed attention) group. This interesting interaction has important implications. It implies, not surprisingly, that a certain level of selectivity is necessary to produce learned inattention. But more importantly, this high level of selectivity is not enough. Learned inattention seems to also require that attention allocation be incrementally learned over time. This suggests that it occurs when people initially consider multiple features, but learn through experience to ignore them (in contrast to a top-down strategy implemented from the beginning to focus on one or few features).

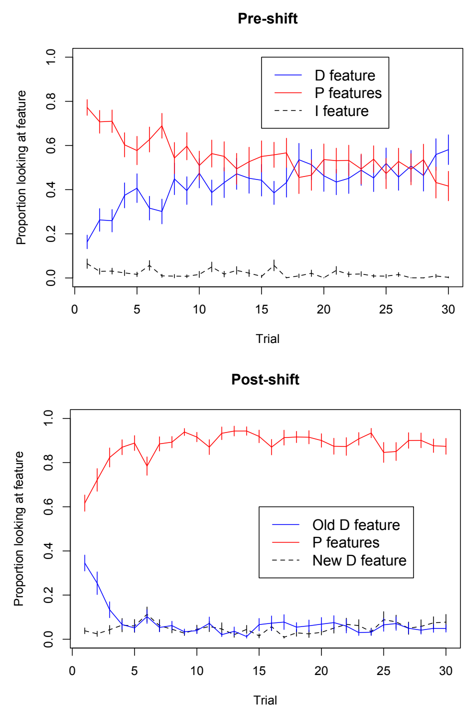

Figure 4. Proportion looking at each feature type during training. Note that the proportion for Probabilistic features is summed across all five Probabilistic features. During initial training looking to the Deterministic feature increased, while looking to the other features decreased (i.e. attention optimization occurred). Post-shift looking to the previously Deterministic feature quickly dropped and was replaced by increased looking to the Probabilistic features, while the previously Irrelevant (now Deterministic feature) remained low. Error bars represent standard error of the mean.

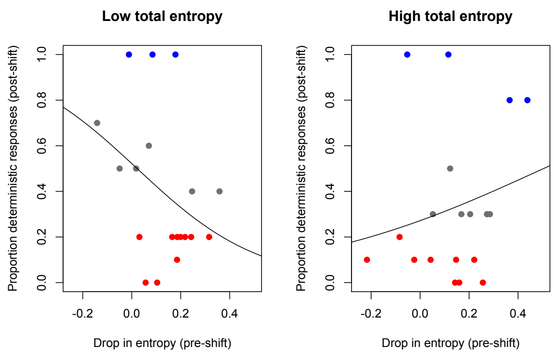

Figure 5. The role of selectivity and attention optimization. Overall entropy (how distributed attention was) interacted with the change in entropy (attention optimization) to produce costs consistent with learned inattention. Participants who were both highly selective (low entropy) and who optimized (large drop in entropy) were the least likely to categorize based on the new deterministic feature after the shift. Line shown the fit line of logistic regression. Dot color indicates deterministic responders (blue), probabilistic responders (red), intermediate (grey).

Discussion

Although selective attention is often effective and efficient, there are potential costs. One particular (longer-term) cost is that selective attention can result in learned inattention to non-selected information, which in turn affects future learning. In the current study, participants performed a category learning task designed to induce learned inattention while we tracked their gaze. Both behavioral and eye-tracking measures showed evidence of learned inattention. When making category judgments after the unexpected shift, participants were less likely to use a stimulus dimension that was previously irrelevant, but was now highly informative. Eye-tracking showed that after the shift occurred, participants quickly shifted attention away from the previously informative (and now irrelevant) feature. Their attention instead shifted to probabilistically predictive features, but they continued to ignore the previously irrelevant, now perfectly predictive, feature—with looking to that feature remaining low and not increasing over the course post-shift training. Participants simply ignored this feature despite its potential usefulness in their task.

The level of learned inattention that was exhibited varied across individuals, though. We used measures of selectivity and attention optimization for each individual to determine what aspects of initial learning best predicted the level of learned inattention that occurred. Our results suggest an interaction between these two measures, such that learned inattention was most likely for participants who were overall highly selective, but importantly, who optimized their attention over time. Participants who were highly selective from the beginning, and so did not optimize attention over time, did not show high levels of learned inattention (see Figure 5). These results suggest that learned inattention crucially depends on incremental learning over time, and is not simply an effect of ignoring sources of information, but of *learning* to inhibit them after initially considering them.

Participants who did optimize attention over time, but whose attention was overall relatively distributed (having high entropy), also did not show high levels of learned inattention. One possibility is that these participants needed more time to optimize attention before reaching the level required for substantial learned inattention to occur, and that with more training trials they would reach that level.

That both the level of selectivity and attention optimization predict individual differences in learned attention has important implications for cognitive development. Young children tend to distribute their attention broadly and do not optimize attention as much as adults do (Best, Yim, & Sloutsky, 2013; Deng & Sloutsky, 2016), so they may be largely protected against the adverse effects of learned inattention.

Allocating attention is always a tradeoff: selective attention results in more efficient processing of attended information, but has several potential pitfalls, including learned inattention. In contrast, if attention is distributed, processing is less efficient, but these traps are avoided. Therefore, to allocate attention effectively estimations must be made about information's potential future relevance. With less general knowledge, children have less basis to make solid conclusions about what might and might not be useful to know later. Additionally, these types of costs could be particularly damaging early in the learning process, and so perhaps children's tendency to distribute attention may be not only a result of immature control, but also adaptive for their particular situation. Understanding the developmental differences in this process is an important direction for future research.

Acknowledgements

Acknowledgments to be completed after review to facilitate double-blind review process.

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological model of multiple systems in category learning. *Psychological Review, 105*, 442–481.

Best, C. A., Yim, H., & Sloutsky, V. M. (2013). The cost of selective attention in category learning: Developmental differences between adults and infants. *Journal of Experimental Child Psychology, 116*(2), 105-119.

Blair, M. R., Watson, M. R., & Meier, K. M. (2009). Errors, efficiency, and the interplay between attention and category learning. *Cognition*, *112*(2), 330-336.

Deng, W. S., & Sloutsky, V. M. (2016). Selective attention, diffused attention, and the development of categorization. *Cognitive Psychology*, *91*, 24-62.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193-222.

Hampton, J. A. (1995). Testing the prototype theory of concepts. *Journal of Memory and Language*, *34*(5), 686-708.

Hanania, R., & Smith, L. B. (2010). Selective attention and attention switching: towards a unified developmental approach. *Developmental Science, 13*(4), 622–635.

Hoffman, A. B., & Rehder, B. (2010). The costs of supervised classification: The effect of learning task on conceptual flexibility. *Journal of Experimental Psychology: General*, *139*, 319–340.

Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review*, *7*(4), 636-645.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111*, 309–332.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological review*, *85*(3), 207.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57.

Pashler, H., Johnston, J. C., & Ruthruff, E. (2001). Attention and performance. *Annual Review of Psychology*, *52*(1), 629-651.

Plebanek, D. J., & Sloutsky, V. M. (2017). Costs of selective attention: when children notice what adults miss. *Psychological Science*, *28*(6), 723-732.

Plude, D. J., Enns, J. T., & Brodeur, D. (1994). The development of selective attention: a life-span overview. *Acta Psychologica*, *86*(2-3), 227-272.

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

Rich, A., & Gureckis, T. (2018). The limits of learning: Exploration, generalization, and the development of learning traps. *Journal of Experimental Psychology: General, 147*(11), 1553-1570.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, 27(3), 379-423.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, *75*(13), 1-41.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1411–1430.

Yantis, S. (2000). Goal-directed and stimulus-driven determinants of attentional control. *Attention and Performance*, *18*, 73–103.