

Rapid Learning in Early Attentional Processing: Bayesian Estimation of Trial-by-Trial Updating

Aaron Cochrane (akcochrane@wisc.edu)
Vanessa Simmering (simmering@wisc.edu)
Joseph L. Austerweil (austerweil@wisc.edu)
C. Shawn Green (cshawn.green@wisc.edu)
Department of Psychology, 1202 W. Johnson Street
Madison, WI 53706 USA

Abstract

All agents must constantly learn from dynamic environments to optimize their behaviors. For instance, it is necessary in new environments to learn how to distribute attention – i.e., which stimuli are relevant, and thus should be selected for greater processing, and which are irrelevant, and should be suppressed. Despite this, many experiments implicitly assume that attentional control is a static process (by averaging performance over large blocks of trials). By developing and utilizing new statistical tools, here we demonstrate that the effect of flanking items on response times to a central item (often utilized as an index of attentional control) is systematically and continuously influenced through time by the statistics of the flanking items. We discuss the implications of this finding from the perspective of examining individual differences – where traditional data analysis approaches may confound the rate at which attentional filtering changes through time with the asymptotic ability to filter.

Keywords: Learning; attention; statistical inference; Bayesian analysis

Introduction

In human, animal, and artificial cognitive architectures, learning to utilize available information for goal-directed behavior is a crucial ability. Critically, in nearly all theoretical models, learning is viewed as an inherently continuous process – with each new data point that is sampled resulting in some concomitant change in knowledge and behavior. One consequence of this is that, even in cases where huge amounts of data have already been sampled (and thus each new data point will change behavior by only a small amount), there remain very few behaviors that would be posited to be fully static or unchanging over time.

Despite this theoretical foundation, in practice the standard analysis approach taken for tasks in the psychological literature implicitly assumes that behavior is in fact static over some period of time (if not the entirety of the task experience). This tendency is even seen in the study of learning – where it has traditionally been quite common to examine performance divided into arbitrary discrete timescales. That is, performance is typically divided into methodologically useful units, such as blocks of training or testing. Data within the block is analyzed under the assumption that the same process generated the data within the entirety of a given block. This generally takes the form of aggregating within-block performance using some function

or algorithm in order to summarize performance, e.g., percent correct (Ahissar & Hochstein, 1997), logistic psychometric function (Schütt, Harmeling, Macke, & Wichmann, 2016), or Drift Diffusion Model (DDM) parameters (White, Brown, & Ratcliff, 2012; Wiecki, Sofer, & Frank, 2013). While such aggregation has some methodological and analytic utility, in terms of its simplification of the data and behavior, it imposes artificial structure upon learning processes that are theoretically independent of that structure.

Not only is it the case that the effects of interest are almost certainly independent of block structures, but the learning that occurs within blocks may itself be theoretically informative. We have previously shown, by developing and employing a time-continuous data analytic approach for assessing visual perceptual learning task performance, that it is possible to differentiate between two distinct forms of learning generalization. The first type of generalization leads to immediate benefits (i.e., is present from the very first trial of a new task), while the second involves no immediate changes in performance on new tasks, but instead new tasks are learned more quickly. These distinct patterns have enormous theoretical importance, as they are generated via completely different mechanisms. Yet they are impossible to differentiate via traditional data analytic techniques that aggregate performance over large blocks of trials (Kattner, Cochrane, Cox, Gorman, & Green, 2017).

Here we extend the general approach to modeling performance as a continuous-function of time to an area where the potential for learning effects are much more rarely considered – the study of attentional control over peripheral (i.e., non-target) processing. Indeed, the analytic techniques utilized in this domain nearly always implicitly assume that performance is static through time. For instance, such aggregation-based analyses are commonly utilized as individual difference metrics, to identify atypical populations (e.g., ADHD; Westerberg, Hirvikoski, Forssberg, & Klingberg, 2004), to characterize development (Rueda et al., 2004), or to simply benchmark difficulties of a test (Edwards et al., 2006). This is despite the fact that it is unlikely to be the case that participants can enter a task with perfect knowledge regarding the spatial and temporal properties of the task-relevant (i.e., to-be-attended target) stimuli or the spatial and temporal properties of distractors (i.e., the to-be-

ignored stimuli). Here we examine the extent to which learning can be identified and modeled in one extremely common index of attentional control – flanker task performance.

Previous work

Within the study of attentional control, certain domains have largely been understood as automatic and independent of the associations or statistics of the environment and thus reasonably impervious to learning (Treisman, 1985; Wolfe, 1994). However, it has also been recognized that in order for a person to interact optimally with their environment, they must constantly weight the utility of the information available to them at all levels of processing. Indeed, dynamic allocation of attention is a core aspect of human ability to interact with the world. Flexibly adapting attention to the changing demands of the environment allows efficient and accurate goal-directed processing of the relevant information available.

When searching for a target in the visual world, distracting items become increasingly easy to suppress as they become increasingly distinct from the target. In the opposite case, when responses between searched-for items and irrelevant items are opposing, a marked increase in response times (i.e., increase in effort needed) to the relevant items is observed. Remarkably, this occurs even when participants are given explicit instructions regarding where and when the relevant item will occur (as well as any irrelevant distracting items). One paradigm in this vein is the arrow flanker task, derived from Eriksen and Eriksen (1974). In this task participants simply press the right keyboard arrow when a central stimulus is a right-pointing arrow, and they press the left keyboard arrow when the central stimulus is a left-pointing arrow. Two other arrows appear on either side of the central arrow pointing in either the same or opposite direction of the central arrow. When the flanking arrows point in the same direction as the central arrow, response times tend to be faster and more accurate than when flankers point in the opposite direction as the central arrow. The differences between congruent-flanker response times and incongruent flanker-response times are largely understood as slowing that occurs due to processing of response-incompatible stimuli, and the magnitude of this difference is often referred to as the "flanker effect."

The flanker effect has been explored in many settings and interpreted in a wide variety of ways. These primarily involve appeals to a neuropsychological executive function or conflict-resolution mechanism (Fan, McCandliss, Sommer, Raz, & Posner, 2002; Machizawa & Driver, 2011). Through this lens the flanker effect has been correlated with such measures as age (Rueda et al., 2004) and cortical thickness (Westlye, Grydeland, Walhovd, & Fjell, 2011). In each of these paradigms, the flanker effect is interpreted as a stable ability within individuals; in effect, it is seen as a robust index of one's ability to control attention and rapidly suppress

distracting information. Unfortunately, this perspective disregards another central aspect of humans' interaction with their environments: The necessity of learning how to weight information appropriately given past experience. While previous research has assumed that psychology tasks (e.g., flanker task) index a constant ability level, we instead posit that learning occurs to some extent (Lehle & Hübner, 2008). This learning in peripheral attention occurs despite the fact that participants are given explicit verbal instructions regarding the time, place, and attributes of the to-be-attended information.

We first demonstrate experimentally, using biased task statistics, that block-level analyses show learning in adult humans' performance on a flanker task. Next, we propose a novel analysis of flanker task response time in which performance is modeled as a function of experience (i.e., trial number). We note that participants were not informed of any learning component to the study. Our analyses show that, even this context, decomposition of performance into parameters of continuous learning reveals the dynamics of humans' interactions with their environments.

Method

Participants and procedure

Forty-seven undergraduate participants from the University of Wisconsin-Madison completed all tasks for course credit. One participant was excluded for missing data. The entire study consisted of three tasks – a flanker task, a Useful Field of View task (UFOV; Ball & Owsley, 1993), and a Multiple Object Tracking task (MOT; Pylyshyn & Storm, 1988). Here, for brevity, we will only consider performance on the flanker task.

The flanker task was modeled after that utilized by Rueda et al (2004). Stimuli were colored fish with arrows overlaid on top pointing in either a leftward or rightward direction. The full flanker task was divided into 5 blocks. Each flanker block included feedback regarding response time and incorrectness. Participants first completed a block of 50 no-flanker trials, second a 250-trial block of either 20%-congruent or 80%-congruent flanker trials (randomly-chosen, with the remaining trials being incongruent), third a 250-trial block of 50%-congruent flanker trials, fourth the biased block that they did not already complete (i.e., 250 trials of either 20%-congruent or 80%-congruent), and fifth a 50-trial block with no flankers. In all cases the participants' task was the same – to indicate the direction of the center fish/arrow as quickly and accurately as possible. Our key questions were whether we would see: (1) differences in performance at a broad scale – in terms of different patterns of response times to the congruent and incongruent trials in the different blocks; and (2) at a continuous time-scale – indicating how such shifts are learned through time.

We note that the other two tasks (UFOV and MOT) were completed between the biased blocks of flanker tasks in order

to reduce monotony and obscure the biasing of the task statistics. For example, by including MOT after the 50% congruent condition and before the 80% congruent condition, we intended that participants would have less carry-over of learning from the 50% congruent condition to the 80% congruent condition.

Analysis

We conducted two Bayesian analyses. The first involved fitting a hierarchical linear model to each block's data. This was designed to test whether, in environments with different statistics, people alter their processing of non-relevant information. The second fit time-evolving weighting parameters to the flanker effect for each plot. This model was able to distinguish which component(s) of learning differ between conditions, as well as demonstrating a novel estimate of continuously changing attentional allocation.

By-block analysis

As an initial demonstration that participants learn to alter their attention in response to changing environmental statistics, a Bayesian multilevel linear model was fit. This model tested the effects of condition compatibility proportion, trial flanker compatibility, and the interaction between these two variables. Block k 's free parameters for subject s , $\beta^{(s)}_{0k}, \dots, \beta^{(s)}_{3k}$ was drawn from a participant-level distribution, which in turn was drawn from a parent distribution shared by all participants.

$$\begin{aligned} \tau_A &\sim G(.001, .001) \\ \beta_i &\sim N(0, 100) \\ \tau^{(s)}_i &\sim G(.001, .001) \\ \beta^{(1)}_i, \dots, \beta^{(s)}_i &| \beta_i \sim N(\beta_i, \tau_A) \\ \beta^{(s)}_{i1}, \dots, \beta^{(s)}_{ik} &| \beta^{(s)}_i \sim N(\beta^{(s)}_i, \tau^{(s)}_i) \\ \log RT &\sim N(\beta^{(s)}_{0k} + \beta^{(s)}_{1k} *congruence \\ &+ \beta^{(s)}_{2k} *percent_congruent \\ &+ \beta^{(s)}_{3k} *congruence *percent_congruent, \tau_A) \end{aligned}$$

where τ_A is a precision parameter for the data distribution and $\tau^{(s)}_i$ is a precision parameter shared across blocks per subject.

This model considered only the last 200 trials in each block in order to characterize asymptotic performance. The predicted outcome of this model was that response times to congruent and incongruent trials would be different from one another (as has been seen in all previous research, with congruent RTs being faster than incongruent RTs), but with these differences themselves differing across varying levels of task statistics, meaning that participants had in fact shifted their behavior based upon the task statistics (i.e., had learned). This response time difference should be evident when controlling for trial congruence as well as individual differences in overall response times. This result would provide evidence that the following analysis, on the time course of learning, would be justified.

By-trial analysis

After testing for block-wise differences between conditions in the magnitude of the flanker effect, we defined a generative process that we hypothesized would give rise to continuous changes in the flanker effect. Fitting parameter estimates to this process would provide hierarchical estimates of the inter-individual and intra-individual variations in the adaptation of attention to environmental statistics.

This analysis assumed two interacting processes. First, that each individual has a stable, domain-general speed-of-processing (SoP) ability that indexes how fast that person can perceive, attend to, and react to their environments (Conway, Cowan, Bunting, Theriault, & Minkoff, 2002). In the flanker task, this would be akin to the response time to the central stimulus when disregarding any effect of the peripheral stimuli. Second, there is a flanker-congruence related offset to the baseline response time. Here that offset is modeled as an additive shift to the baseline on a log scale, which translates to a multiplicative shift in raw response times. Approaches to flanker analysis that equate congruent-flanker trials with no-flanker trials would parameterize this relation as simply an additive component to the baseline. However, in order to remain sensitive to the possibility of the shift adding to the response time in the incongruent-flanker condition while subtracting from the response time in the congruent-flanker condition (i.e., speeding), here the shift is parameterized as symmetrically adding or subtracting to a central baseline log SoP response time ability. That is, we maintain the possibility that participants use congruent flankers to speed up their response times (noting though that the high-level pattern of results with respect to learning should not be strongly dependent on this choice).

Typical individual-differences flanker analyses utilized in the field assume that the congruency-related shift (whether solely positive or not) is stable across the course of the flanker task. Indeed, in order to remain valid, the shift must even be constant across several repetitions the task by a single person. Here that assumption is relaxed. Rather than assume a constant additive shift due to flanker type, the additive shift is assumed to be learned. That is, participants update their attention to flanking items (i.e., their additive shift) throughout the task in response to the utility of attending to peripheral items. Here the additive shift is parameterized as exponential decay as a function of trial number, $(a+b*c^{-t})$. Exponential learning functions are extremely common in many fields, and provide concise characterizations of the time course of learning (Heathcote, Brown, & Mewhort, 2000).

Thus, the generative process assumes the response time on a given trial is the following:

$$\begin{aligned} \beta_A &\sim N(-1, .1) & \tau_A &\sim G(100, .1) \\ \beta_{\text{asym}} &\sim N(.1, .01) & \beta_{\text{scale}} &\sim N(.01, .1) \\ \beta_{\text{rate}} &\sim N(1.1, .1) \end{aligned}$$

$$\begin{aligned} \beta^{(1)}_{\text{SoP}}, \dots, \beta^{(s)}_{\text{SoP}} &| \beta_{\text{SoP}} \sim N(\beta_A, \tau_A) \\ \beta^{(1)}_{\text{asym}}, \dots, \beta^{(s)}_{\text{asym}} &| \beta_{\text{asym}} \sim N(\beta_{\text{asym}}, \tau_A) T(0, 2) \\ \beta^{(1)}_{\text{scale}}, \dots, \beta^{(s)}_{\text{scale}} &| \beta_{\text{scale}} \sim N(\beta_{\text{scale}}, \tau_A) T(0, 2) \\ \beta^{(1)}_{\text{rate}}, \dots, \beta^{(s)}_{\text{rate}} &| \beta_{\text{rate}} \sim N(\beta_{\text{rate}}, \tau_A) T(-2, 2) \end{aligned}$$

$$\begin{aligned} \beta^{(s)}_{\text{asym}(1)}, \dots, \beta^{(s)}_{\text{asym}(k)} &| \beta^{(s)}_{\text{asym}} \sim N(\beta^{(s)}_{\text{asym}}, \tau_A) \\ \beta^{(s)}_{\text{scale}(1)}, \dots, \beta^{(s)}_{\text{scale}(k)} &| \beta^{(s)}_{\text{scale}} \sim N(\beta^{(s)}_{\text{scale}}, \tau_A) \\ \beta^{(s)}_{\text{rate}(1)}, \dots, \beta^{(s)}_{\text{rate}(k)} &| \beta^{(s)}_{\text{rate}} \sim N(\beta^{(s)}_{\text{rate}}, \tau_A) \end{aligned}$$

$$\log RT \sim N(\beta^{(s)}_{\text{SoP}} + \text{congruence} * (\beta^{(s)}_{\text{asym}(k)} + \beta^{(s)}_{\text{scale}(k)} * \beta^{(s)}_{\text{rate}(k)}^{-\text{trial}}), \tau_A)$$

where $T(a,b)$ truncates a distribution to the range (a,b) .

Each of the three learning parameters of interest (flanker offset asymptote, scaling and exponent terms) were estimated as normal distributions for each block, with the mean of this normal being drawn from participant-level asymptote (truncated at 0 and 2), scaling (truncated at -2 and 2), and exponent (rate; truncated at 0 and 2) normal distributions. Truncations were imposed at values beyond which model behavior would be qualitatively very different than the theoretical generative model. In particular, the entire peripheral attention term should evaluate to less than 1 in every instance in order to be a sensible fit to the data-generation process (i.e., the difference between incongruent-flanker trials and congruent-flanker trials is never more than 2 in log-RT space).

All other prior distributions were non-truncated, with normal priors for all mu distributions and gamma priors for all gamma distributions and precision distributions. Given the primary interest in comparing between-block within-subjects variation as a function of block statistics, all variation of interest should be caused by the data and not by prior specification.

Results

Bayesian analysis using JAGS implemented in R (Plummer, 2003) was used for parameter approximation. Four chains were burned in for 20,000 samples, then 200,000 samples were drawn for further analysis.

All response time measures were first trimmed to exclude values above 2 seconds and below .05 seconds (120 trials total rejected), as response times outside these bounds are clearly not arising from the processes of interest in this study. In addition, all trials with incorrect responses were excluded (8.9%); further analysis of this incorrect-trial data may be relevant to the core questions of this study, but analysis of this variable was outside the scope of the current paper (see

Limitations section below). After this trimming, the remaining 25,064 response times were log-transformed to better approximate normality. Given this, log-transformed response times varied from -2.99 to 0.66 ($m = -.99$, $sd = .237$).

Convergence

Bayesian analysis appeared to converge in both models. Visual inspection of trace plots, autocorrelation plots, and Gelman-Rubin plots indicated convergence for the majority of estimated parameters. Five parameters in the by-trial analysis, all of which were block-level estimates of the rate parameter (i.e., $\beta^{(s)}_{\text{rate}(k)}$), presented clearly problematic traces and autocorrelations. We excluded the five participants with problematic rate parameter traces from the following by-trial analyses, leaving data from 41 participants.

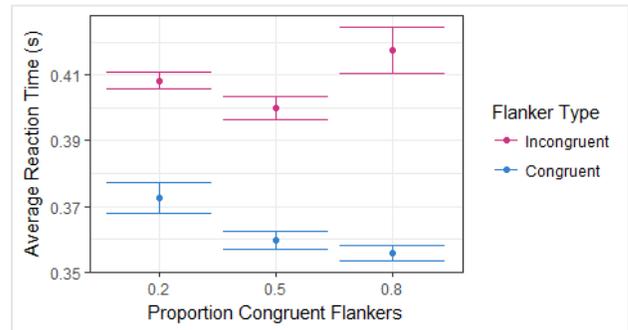


Figure 1. Mean response times, separated by flanker congruence and proportion congruent flankers. Error bars denote 95% confidence intervals across all trials for all participants.

By-block fits

Point estimates of the parameters of interest (i.e., the means of the level-one regression predictors) each provide support for the hypothesis that attention changes with environmental statistics. The 95% credible intervals of these parameters follow the same pattern. This is the case for the effects of flanker congruence (mean= -.081, lower= -.100, upper = -.062), proportion congruence (mean= -.052, lower= -.114, upper = .009), and the interaction between the two (mean= -.046, lower= -.091, upper = -.002). For each of these parameters, the vast majority of the mass of the distribution is on one side of zero. It is evident in Figure 1 that this pattern supports the hypothesized effects; in situations where the participant sees mostly facilitative non-targets, they are faster in responding to congruent-flanker trials while also being slower on incongruent-flanker trials in these situations. In essence, when most of the flankers are congruent, there is an advantage in reducing the extent to which these flankers are filtered. This produces faster RTs on congruent trials, but then causes disproportionate slowing on incongruent trials. Meanwhile, in conditions where most of the trials are incongruent, there is virtue in strongly filtering all flankers. Note a lesser facilitatory effect is then seen for the congruent

trials, but the magnitude of the drop-off in RT on incongruent trials is reduced.

Although we explore this in greater detail below, the data already indicates that such behavioral shifts must be learned over many trials. For instance, when subsets of the 50-50 condition are analyzed – either short runs in which four congruent trials were followed by an incongruent trial, or four incongruent trials were followed by a congruent trial (effectively creating miniature “80-20” or “20-80” conditions) the change in response time from the fourth to the fifth trials is no different than the overall differences between congruent and incongruent trials in the 50-50 condition (both $|t| < 1.1$). This suggests that the large-scale differences seen across the different blocks are the result of a longer-term learning process. The following by-trial analysis further tested the time course of learning.

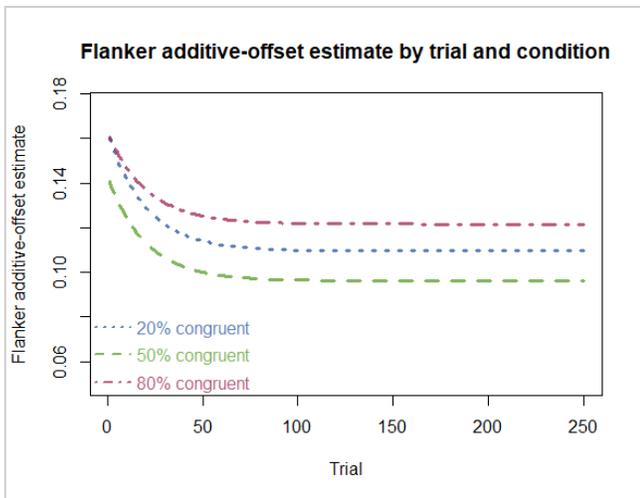


Figure 2. Mean by-trial half-flanker effect (i.e., the additive component of the flanker effect, which would have a subtractive component mirrored across $y=0$). Separated by block type (percent of congruent trials).

By-trial fits

Means of evaluated fits are shown in Figure 2. The three main parameters were asymptote, which indicated the flanker effect size after learning, scale, which indicated the magnitude and direction of learning, and rate, which indicated the relative speed of learning. We conducted preliminary frequentist nonparametric comparisons of fit parameters in the extreme conditions via within-subjects Wilcoxon signed-rank tests that compared the 20% congruent to the 80% congruent conditions. The difference between the rate parameters in 20%-congruent condition (median = 1.0517) and 80%-congruent condition (median = 1.0502) was 0.0015, $V = 536$, $p = 0.176$. The difference between the asymptote parameters in 20%-congruent condition (median = 0.109) and 80%-congruent condition (median = 0.129) was 0.019, $V = 301$, $p = 0.095$. The difference between the scale

parameters in 20%-congruent condition (median = 0.051) and 80%-congruent condition (median = 0.041) was 0.010, $V = 771$, $p < .001$. These analyses provide preliminary evidence for reliable differences in learning scale, but not the other parameters, between conditions. Scale determines magnitude as well as direction of learning, making it a reasonable parameter to expect to differ between conditions if we believe that people truly are learning to behave differently.

We next tested three frequentist multilevel models using R packages *lme4* and *pbkrtest*, one for linear changes in each of the three exponential parameters due to changes in congruency statistics.

$$lmer(\text{param} \sim \text{propCongruent} + (\text{propCongruent} | \text{subject}))$$

Each model used the proportion of congruent trials in a block to predict the fit parameter value, while controlling for participant-level random effects. The rate parameter was not reliably predicted by task statistics, $F(1,44.7) = .817$, $p > .35$. The asymptote likewise fell short of conventional statistical significance, $F(1,43.1) = 3.97$, $p = .053$. In contrast, but in concurrence with the Wilcoxon test reported above, the scale parameter was linearly predicted by varying flanker-congruency proportions, $F(1,44.0) = 26.91$, $p < .001$.

Limitations

These analyses have certain weaknesses and shortcomings. For example, the apparent nonmonotonicity of fits with regard to task statistics could be an artifact of block order effects that indicates learning to learn. In addition, this work, meant as a preliminary demonstration, utilizes a simplistic measure of flanker performance. Response times for a given trial are assumed to be additively shifted from a baseline (in a log-transformed scale), while trials with incorrect responses are omitted. Further work should explore continuous learning-related changes via models that capture both response time and accuracy (as in Drift Diffusion Models - DDMs), as these are more meaningful decompositions of performance than the only-correct log-transformed RTs reported here. Doing so may require longer learning blocks, as DDMs with relatively high numbers of parameters are unlikely to recover reliable estimates of learning parameters given a mere 250 learning trials per condition. While hierarchical modelling would somewhat alleviate these concerns by providing stability (i.e., lower-level parameters could only be estimated from the distributions of higher-level parameters), we refrained from testing these models with high numbers of free parameters here. One direction for future work could be to apply hierarchical DDM parameters with covariates (Wiecki et al., 2013) to the problem of trial-by-trial learning by specifying a functional form (e.g., exponential decay). Many other additional parameters could be fit as well, such as changing SoP values or asymmetric flanker effects (e.g., additive effects due to incongruent trials being larger than subtractive effects due to congruent trials).

Discussion

Here we demonstrate that bottom-up attention is reliably influenced by environmental statistics. That is, the degree of filtering demonstrated by a participant is a function of some amount of learning – rather than fully reflecting a static ability. We provide evidence for a quantitative dissociation between the dynamics of learning and the stable individual differences that interact to give rise to the overall pattern of behavior in the flanker task. The scale (indicating size and/or direction) of learning is clearly changed by environmental statistics. The asymptote appears to be changed as well, although our data indicates that this change is not linear or even monotonic (see Figure 2). While many questions remain to be examined regarding learning in attentional tasks, this first step provides impetus to further address how to best quantitatively decompose behavior in single tasks into separate processes, including a learning process.

The key implication of this work is thus that individual differences approaches to attention, and cognition more generally, would benefit from integrating analyses sensitive to the effects of learning. Performance differences between individuals due to learning may be mechanistically distinct from individual differences arising from, for example, stable differences in distractor suppression. Furthermore, such considerations may shed light on other areas of the field – for instance, in examining test-retest reliabilities. Indeed, the test-retest reliability of flanker tasks has typically not been found to be high. However, this may be due to failing to account for learning from test to test. By implementing experimental paradigms and analytical methods capable of identifying the relative contributions of these processes, further light may be shed on the mechanistic underpinnings of a wide array of typical processing (e.g., fluid intelligence) as well as atypical (e.g., ADHD, anxiety).

References

- Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*(6631), 401–406.
- Ball, K., & Owsley, C. (1993). The useful field of view test: a new technique for evaluating age-related declines in visual function. *Journal of the American Optometric Association*, *64*(1), 71–79.
- Conway, A. R., Cowan, N., Bunting, M. F., Theriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163–183.
- Edwards, J., Ross, L., Wadley, V., Clay, O., Crowe, M., Roenker, D., & Ball, K. (2006). The useful field of view test: Normative data for older adults. *Archives of Clinical Neuropsychology*, *21*(4), 275–286.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*, 143–149.
- Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, *14*(3), 340–347.
- Heathcote, A., Brown, S., & Mewhort, D. J. (2000). The power law repealed: the case for an exponential law of practice. *Psychonomic Bulletin & Review*, *7*(2), 185–207.
- Kattner, F., Cochrane, A., Cox, C. R., Gorman, T. E., & Green, C. S. (2017). Perceptual Learning Generalization from Sequential Perceptual Training as a Change in Learning Rate. *Current Biology*, *27*(6), 840–846.
- Lehle, C., & Hübner, R. (2008). On-the-fly adaptation of selectivity in the flanker task. *Psychonomic Bulletin & Review*, *15*(4), 814–818.
- Machizawa, M. G., & Driver, J. (2011). Principal component analysis of behavioural individual differences suggests that particular aspects of visual working memory may relate to specific aspects of attention. *Neuropsychologia*, *49*(6), 1518–1526.
- Plummer, M. (2003). JAGS: A Program for Analysis of Bayesian Graphical Models using Gibbs Sampling. *3rd International Workshop on Distributed Statistical Computing*, 124.
- Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, *3*(3), 179–197.
- Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia*, *42*(8), 1029–1040.
- Schütt, H. H., Harmeling, S., Macke, J. H., & Wichmann, F. A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Research*, *122*, 105–123.
- Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing*, *31*, 156–177.
- Westerberg, H., Hirvikoski, T., Forssberg, H., & Klingberg, T. (2004). Visuo-spatial working memory span: a sensitive measure of cognitive deficits in children with ADHD. *Child Neuropsychology*, *10*(3), 155–161.
- Westlye, L. T., Grydeland, H., Walhovd, K. B., & Fjell, A. M. (2011). Associations between Regional Cortical Thickness and Attentional Networks as Measured by the Attention Network Test. *Cerebral Cortex*, *21*(2), 345–356.
- White, C. N., Brown, S., & Ratcliff, R. (2012). A test of Bayesian observer models of processing in the Eriksen flanker task. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 489–497.
- Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, *7*.
- Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*, 202–238.