



Modulation of compatibility effects in response to experience: Two tests of initial and sequential learning

Aaron Cochrane¹ · Vanessa Simmering² · C. Shawn Green¹

Accepted: 14 October 2020 / Published online: 9 November 2020
© The Psychonomic Society, Inc. 2020

Abstract

Attentional control is a key component of goal-directed behavior. Modulation of this control in response to the statistics of the environment allows for flexible processing or suppression of relevant and irrelevant items in the environment. Modulation occurs robustly in compatibility-based attentional tasks, where incompatibility-related slowing is reduced when incompatible events are likely (i.e., the proportion compatibility effect; PCE). The PCE implicates dynamic changes in the measured compatibility effects that are central to fields of study such as attention, executive functions, and cognitive control. In these fields, stability in compatibility effects are generally assumed, which may be problematic if individual or group differences in measured compatibility effects may arise from differences in statistical learning speed or magnitude. Further, the sequential nature of many studies may lead the learning of certain statistics to be inadvertently applied to future behaviors. Here, we report tests of learning the PCE across conditions of task statistics and sequential blocks. We then test for the influence of feedback on the development of the PCE. We find clear evidence for the PCE, but no conclusive evidence for its slow development through experience. Initial experience with more incompatible trials selectively mitigated performance decreases in a subsequent block. Despite the lack of behavioral changes associated with patterns of learning, systematic within-task changes in compatibility effects remain an important possible source of variation in a wide range of attention research.

Keywords Cognitive and attentional control · Attention in learning · Attention: Selective

Introduction

In a seminal paper, B. A. Ericksen and Ericksen (1974) demonstrated the flexibility and ubiquity of response competition as a window into visual attention. Critically, the tasks that were developed in this vein of research frequently necessitated that participants learn arbitrary stimulus–response mappings (e.g., an *S* stimulus associated with a right index finger button press). Accordingly, the associated response competition effects that were of primary importance as dependent variables in these tasks were likewise necessarily the products of learning. Indeed, participants did not enter tasks utilizing arbitrary

stimulus–response mappings already having reaction times that were slower when a particular button needed to be pressed in the context of a particular stimulus, as compared with when the same button needed to be pressed in the context of some other stimulus. Instead, these patterns had to emerge as the participants internalized the relationship between the various stimuli and button presses (i.e., response competition). An additional dimension of task demands must be learned as well. Even in contexts wherein prepotent stimulus–response mappings exist (e.g., pressing a right arrow key when a right-pointing arrow is displayed on a screen), participants must learn through experience the statistical distributions of the stimuli and associated responses.

This perspective aligns with the overarching idea that the behavioral markers of attentional phenomena often require certain types of experience to be manifested. This is true, for instance, of the proportion compatibility effect (PCE), which is observed when participants experience unequal numbers of trials that include compatible (response-congruent) or incompatible (response-incongruent) distractors. More specifically, in the PCE, experimental conditions that involve more frequent response competition (i.e., a greater amount or degree of

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13414-020-02181-1>.

✉ Aaron Cochrane
akcochrane@wisc.edu

¹ Department of Psychology, University of Wisconsin–Madison, 1202 W. Johnson St, Madison, WI 53706, USA

² Mosaic Consulting LLC, Madison, WI, USA

incompatible responses between targets and distractors) are associated with a reduction in response time (RT) differences due to compatibility effects (Braem et al., 2019; Gratton, Coles, & Donchin, 1992; Logan & Zbrodoff, 1979).

Several potential sources of this pattern have been proposed. One possibility is that repeated experiences with response conflict causes an augmentation of attentional control (Bugg & Crump, 2012; Gratton et al., 1992; Lehle & Hübner, 2008). This increase in attentional control would decrease the influence of interfering distractors and in turn produce a convergence in reaction times for compatible and incompatible trials. In the limit, if there was a perfect attentional filtering process, and thus no processing of interfering distractors, RTs should be the same regardless of whether the distractors were response-compatible or response-incompatible (i.e., it would be as if the distractors were “not there”). A second set of proposals has characterized the PCE as arising from an interaction between task difficulty and low-level learning processes (Abrahamse, Braem, Notebaert, & Verguts, 2016; Schmidt, 2016). Broadly, according to one version of this perspective, more-frequent experiences (i.e., incompatible or compatible trials) are learned most and thus experience a disproportionate decrease in RT. However, the underlying difficulty of incompatible trials nevertheless causes RTs for these trials to remain higher than compatible trials, regardless of experience. This theory therefore also predicts a convergence in compatible and incompatible RTs when incompatible trials are frequent. Finally, a third possible explanation for the PCE notes that while experience with multiple trials is clearly necessary to observe the PCE, this experience does not necessarily need to be on a time scale over which the statistics of the task could be adapted to or learned. Instead, short-range trial-to-trial interactions could cause brief variations in behavior that compound to demonstrate the PCE (e.g., the congruency sequence effect; Gratton et al., 1992). Importantly, though, short-term effects such as priming are not exclusive of longer-term learning or cognitive control modulation in response to task statistics, and in practice the mechanisms are likely to interact (Davelaar & Stevens, 2009).

While the models above were primarily developed to capture PCE effects that emerged within a short time scale (e.g., within a block of trials or even within pairs of trials), another implication of a learning-centered perspective on the PCE is the possibility of even longer-term dependencies. One such possibility is that earlier experience with an attentionally demanding task may implicitly teach certain patterns of attentional modulation, rapid inference regarding stimuli and their statistical characteristics, expectations regarding motor responses, or other context-bound information relevant to successful task performance (Schmidt & Weissman, 2014). Longer-term dependencies may not be constrained to just

stimulus–response mappings or short-term attentional modulations; future learning may itself be altered by the previous learning environment (Kattner, Cochrane, Cox, Gorman, & Green, 2017). That is, initial task experience may implicitly teach participants something about the global context which has longer-term effects on response times for compatible or incompatible trials.

Consistent with these ideas, block-to-block carryover effects have been observed in PCE. For example, using a Stroop task, Abrahamse, Duthoo, Notebaert, and Risko (2013) found that majority-compatible blocks were associated with much larger compatibility effects when they were the first block experienced than when preceded by a majority-incompatible block. Meanwhile, the magnitude of compatibility effects in majority-incompatible blocks was similar regardless of prior experience. Abrahamse et al. (2013) attributed these asymmetrical shifts between low-to-high versus high-to-low proportions of compatible experiences to attentional modulation that unfolded over the course of hundreds of trials. Specifically, when participants’ first experience was with a majority-incompatible block, this resulted in a change of attentional focus that not only reduced the magnitude of the compatibility effect within that block (i.e., as would be seen in the typical PCE), but that persisted into the next block (thereby reducing the magnitude of the compatibility effect in that next block as well). Critically, under certain conditions, it was posited that performance shifted too slowly to be captured within their experimental time scale (i.e., the 240 or 288 trials that were utilized). Furthermore, averaging across all RTs within experimental blocks precluded inferences regarding within-block change. These methods and interpretations stand in contrast with the possibility that statistical learning and attentional modulations may occur very rapidly (e.g., trial-to-trial changes in compatibility effects; Braem et al., 2019; Gratton et al., 1992). Only by measuring change on shorter time scales could the possibility of rapid adaptation and learning be tested.

Importantly, the need to consider the possibility of time-evolving processes in tasks where response competition effects are the primary dependent variables of interest reaches far beyond theoretical questions regarding statistical learning or adaptation of attention. Response competition measures have been utilized in studies ranging from individual differences in cortical anatomy (Westlye, Grydeland, Walhovd, & Fjell, 2011) to cognitive training (Rueda, Rothbart, McCandliss, Saccomanno, & Posner, 2005) and the effects of psychoactive substances (Bailey et al., 2016), and similar measures are central to influential theories of attentional networks (Fan, McCandliss, Sommer, Raz, & Posner, 2002; Petersen & Posner, 2012; Posner & Petersen, 1990) and executive functions (Diamond, 2013; Miyake, 2000). In each of these domains, aggregate measures of response competition have been used as an index of the effectiveness of low-level control and

selection processes, yet the extent to which these processes are being contextually adapted is rarely explicitly examined.

There are several possible, but not well-examined, implications of the PCE for these broader research domains that make use of response competition measures. As noted above, the PCE is a context-dependent change in response competition measures (i.e., the PCE must necessarily unfold over time as an interaction between person and environment because statistical differences in the number of compatible versus incompatible trials only emerge through time). It is unclear, though, the extent to which between-participant variation in response competition, which is of interest to many areas of cognitive psychology, arises due to between-participant differences in (1) stable trait-like abilities or (2) magnitude or rate of adaptation to a given context. Further, attentional modulations and statistical learning may influence participants on time scales reaching beyond single blocks of trials, leading to complex sequential effects in the measurement of response competition. Specifically, cognitive theories and empirical results interpreting flanker-task response competition as a static quantity may instead be inadvertently observing and interpreting varying rates of change through time.

One complication in linking the broader literature that has made use of response competition measures to PCE research is that one core aspect of the task design—the presence or absence of feedback—tends to differ across these domains. Indeed, many research domains have employed task versions that do not include explicit feedback (B. A. Eriksen & Eriksen, 1974; Fan et al., 2002; Miyake, 2000). In many cases, whether implicitly or explicitly, this methodological decision may arise from the assumption that behavior is more stable in the absence of additional signals from the environment (i.e., it could reduce the extent to which participants learn via simple experience with the task itself). In contrast, PCE effects have most commonly been observed in situations where participants are provided with informative feedback (i.e., regarding response time, accuracy, or both; Gratton et al., 1992; Schmidt & Weissman, 2014; Wenke, De Houwer, De Winne, & Liefoghe, 2015). Even when comparing explicit instruction-based learning to lower-level statistical learning, Wenke et al. (2015) included feedback in all experimental conditions. While PCE effects have been occasionally reported in the absence of explicit feedback (e.g., Logan & Zbrodoff, 1979), a time-evolving account of these attentional modulations has not been investigated.

Interactions between feedback and learning or adaptation in the PCE may have one of several consequences. Depending on the role of feedback regarding response times in attentional modulations, feedback may either increase learning and thus differences between conditions (i.e., by strengthening the signals participants

use to alter their performance; Seitz & Dinse, 2007) or decrease learning (and differences between conditions; i.e., by providing an alternative stimulus, other than the target, that decreases the influences of response compatibility stimuli on attentional allocation). As such, independent tests of learning may provide clarity regarding the role of feedback in enhancing or inhibiting adaptive attentional modulations.

Here, we examined several issues regarding the possible time-evolving nature of flanker effects and the PCE. First, because arrow-based flanker tasks are very common in the broader literature on attentional control (Bailey et al., 2016; Davelaar & Stevens, 2009; Fan et al., 2002; Rueda et al., 2005; Sidarus, Palminteri, & Chambon, 2019), while PCE effects have typically been primarily investigated in Stroop-like or Simon-like tasks, we first further confirmed that the canonical PCE was observed in participants' initial block of an arrow-flanker task (with different participants receiving 20%, 50%, or 80% of compatible trials in this first block). Tasks utilizing arrows are an interesting testbed in this case as the stimulus–response mappings are likely largely already in place when individuals come to the lab (i.e., the mapping between a leftward facing arrow and the left arrow key on a keyboard is likely very natural for young adults in the United States), leaving the biased task statistics (more or less compatible/incompatible distractor trials) the main to-be-learned aspect of the task.

Second, by modeling RT change as a function of time, we examined whether we could find explicit evidence of the PCE emerging within this first block of trials. Then, by having participants complete a second block of the flanker task with a different proportion of compatible trials than they had experienced during their first block, we assessed whether, after controlling for the second block's proportion compatible, the change in RTs evident in the second block is systematically related to the first block's proportion compatible (i.e., whether there is carryover). Such a manipulation extends the work of Abrahamse et al. (2013), by examining the possibility of block-to-block carryover at a much finer time scale (as well as in a different task; arrow-based flanker as opposed to Stroop). In this vein, we also examined the extent to which purely local (i.e., trial-to-trial) interactions, rather than long-range (i.e., over the course of full blocks of trials) could explain the RT differences. Finally, given that the PCE is inherently a manifestation of experience, and that the presence/absence of feedback is likely to modulate the impact of long-term experience (e.g., learning), in Experiment 1 we investigated the issues above in the context of the methodological approach (no feedback) that is perhaps more common in broader attention and cognition research. In Experiment 2 we then included

feedback into our design to more closely align with typical studies of the PCE.

Overarching methodological approach across Experiments 1 and 2

Overview of procedures and sample

Participants (Experiment 1: $n = 54$; Experiment 2: $n = 60$; gender: 60.8% female; race: 21.6% Asian; 62.2% White; 16.2% other or multiple) were recruited from an introduction to psychology participant pool and given course credit for their participation. All participants provided informed consent, and all procedures were approved by the University of Wisconsin–Madison Institutional Review Board.

Details of task

Participants completed two blocks of an arrow-flanker task (B. A. Eriksen & Eriksen, 1974; Fan et al., 2002; Rueda et al., 2005). Each block consisted of 400 trials in which each trial could be compatible (i.e., target facing the same direction as flankers) or incompatible (i.e., target facing the opposite direction as flankers; see Fig. S1 in the Supplementary Information). Each block included one of three possible predetermined ratios of compatible to incompatible trials: 20–80, 50–50, or 80–20. Participants were pseudorandomly assigned to conditions such that each participant completed blocks containing two different compatibility proportions.

The flanker task was run in Python using the PsychoPy library on a 22-inch Dell monitor in a dimly lit room. Participants sat approximately 60 cm from the monitor. Stimuli consisted of arrows overlaid on cartoon fish (Cochrane, Simmering, & Green, 2019; Rueda et al., 2005) 1.5 degrees of visual angle wide and placed with centers 1.65 degrees apart (i.e., .15 degrees separating stimuli). The center fish was always presented at the center of the screen. Each block included stimuli of a single randomly chosen color of light purple, green, or orange (see [Supplemental Information](#) for stimuli). A 100-millisecond centrally located cross cue was first presented, after which there was a blank screen for a random time between 100 and 300 milliseconds prior to stimulus onset. Responses were recorded on a standard keyboard by pressing the arrow key corresponding to the target stimulus (i.e., left or right). After response, an 850-ms delay occurred.

Analyses

Incorrect trials were first excluded (5.22% across all participants). Because error rates were very low, no analysis of errors was conducted. Trials with RT over 1.5 seconds (0.46%) or below .2 seconds (0.58%) were then excluded. Where

appropriate, we used linear mixed-effects models, with participant-level random intercepts, using the R package lme4 (Bates, Mächler, Bolker, & Walker, 2015, p. 4) with degrees-of-freedom approximation using the package pbkrtest (Halekoh & Højsgaard, 2014). Proportion compatible was treated as a three-level categorical variable, allowing for asymmetric effects of proportions above or below 50%. In tests of the overall PCE, the reference proportion was set to be the 20% condition to simplify our confirmation of the monotonically increasing flanker effect with increasing proportion of compatible flankers. In all other tests, the reference proportion was set to be the 50% condition, providing for a clear interpretation of possible asymmetric effects. Mixed-effects model coefficients therefore indicated the magnitude of the RT differences between levels of the predictors. We reported overall PCE results as interactions between current-trial block proportion compatible. When testing changes in the flanker effect we first averaged, the RT for each participant's compatible and incompatible trials separately for the first 50 trials and last 50 trials of each block. This provided us with eight mean RTs per participant for subsequent analyses. Change in mean RT was next calculated for each participant's blocks by subtracting compatible trials from incompatible trials (i.e., flanker effect) and subtracting the final-50-trial flanker effect from the first-50-trial flanker effect. Results report differences in these change scores.

Experiment 1 results

Was the canonical PCE observed?

The PCE is typically observed as an increase in the magnitude of the flanker effect in conditions with more compatible trials (or likewise, a decrease in flanker effect in conditions with more incompatible trials). We first tested for this pattern across all data to provide estimates averaging over proportions compatible and block number. We fit a linear mixed-effects model predicting RT with main effects and the interaction of

Table 1 Entire-block proportion flanker effect on compatible and incompatible trials

	Estimate (s)	SE	T
Intercept	0.439	0.010	44.095
50%_compat	−0.012	0.003	−4.161
80%_compat	−0.012	0.003	−4.070
IsIncompatible	0.040	0.003	15.269
50%_compat:IsIncompatible	0.008	0.004	2.344
80%_compat:IsIncompatible	0.018	0.004	4.828

trial incompatibility and proportion compatible in a block, using 20% compatible as a reference, while controlling for the random effect of participant-level mean RT (see Table 1). The predicted PCE would manifest as positive coefficients for the interaction between trial type and higher proportions compatible (i.e., larger incompatibility effects), with possible faster RT on compatible trials taking the form of negative main effect coefficients for higher proportions compatible.

Consistent with the PCE, on average, relative to a block with 20%-compatible trials, blocks with more compatible trials had reliably lower compatible-trial RTs, by around .012 s, and similar incompatible-trial RTs, (and thus a larger flanker effect in the form of reliable interactions; see Fig. 1). In examining the fit model coefficients, the effect on compatible trials appeared to have saturated by 50% compatible, while the effect on incompatible trials continued to increase through 80% compatible.

Does the flanker effect reliably change over the course of participants’ first 400 trials?

As noted in the Introduction, when stimulus–response pairs are arbitrary (e.g., if the two possible targets are a square and a diamond, indicated by pressing the “Z” or the “M” buttons on the keyboard), the flanker effect would seemingly require some degree of learning (i.e., to learn that some stimulus–response pairs are “incompatible” requires an understanding of the task). Yet, in the case of an arrow-based task with natural key

mappings, it is less clear whether task-based experience is necessary to observe the flanker effect. Indeed, university students presumably enter such a task with a great deal of experience, with the idea of left/right arrows as incongruent/opposite one another. Furthermore, as noted in the introduction, the methodological approach used in Experiment 1 (utilizing no feedback) is often employed with the goal of eliminating, or at least reducing, change through time (e.g., learning has the potential to be particularly problematic in research that makes use of the same task across multiple time points).

Despite these caveats, to examine whether we could detect systematic change over the course of participants’ first 400 trials, we tested change in incompatible and compatible RTs using a linear mixed-effects model, predicting change in RT with a fixed effect of trial compatibility and a random intercept for each participant. In this model, participants’ flanker effect within their first block did change reliably. However, interestingly, this change was not in the direction that would be expected from a canonical learning effect (e.g., where RTs become faster through time and flanker effects decrease). Instead, the flanker effect reliably increased by .016s rather than decreased through time ($b = 0.016, T = 2.283, p = .027$). When analyzed separately, reliable change was evident in both incompatible trials ($b = 0.05, T = 4.02, p < .001$) and compatible trials ($b = 0.034, T = 3.16, p = .003$), with the increase in flanker effect noted above being indicative of the further significant increase in incompatible RT relative to compatible RT. Therefore, the results

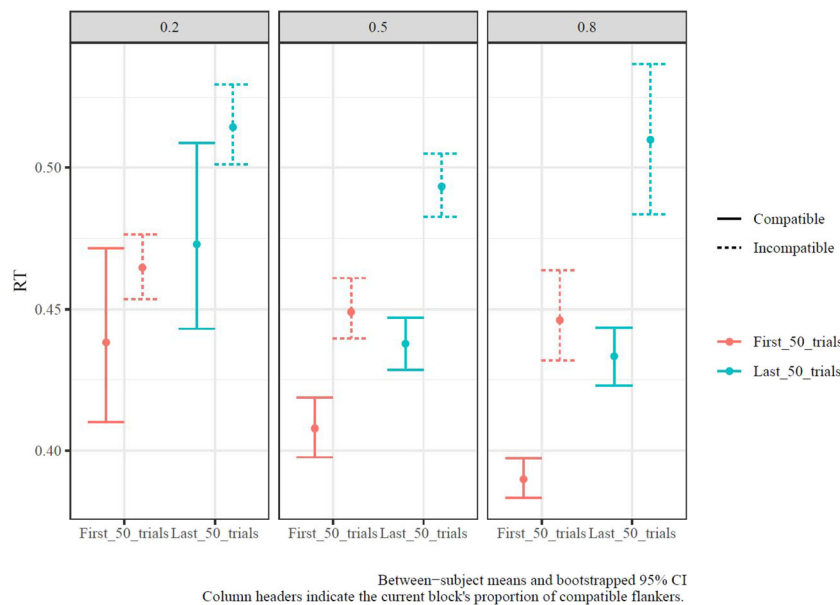


Fig. 1 Block 1 RT in the absence of feedback, separated by the first 50 and last 50 trials. Column headers indicate current block’s proportion compatible. Several outcomes are evident in this plot. First, the expected PCE is seen (i.e., the difference between RTs for compatible and incompatible trials is smaller in the 20%-compatible condition as

compared with the 80%-compatible condition). Second, RTs through time generally increase (i.e., the difference between RTs during the first 50 trials and the last 50 trials). Between-subjects means and 95% CI are indicated. See Fig. S2 in the Supplementary Information for full pattern of results across both blocks and experiments

were inconsistent with the proposition that flanker effects are stable through a block when highly familiar stimuli (arrows) are utilized without feedback. While we observed reliable changes in the flanker effect in participants' first 400 trials, these changes were also not in the direction that would be expected from task learning (i.e., becoming faster through time), nor were the flanker effects clearly optimized over the course of the blocks through statistical learning and adapted attentional control (i.e., smaller flanker effects).

Does compatibility proportion influence within-block changes in response time?

In our first two analyses, we observed canonical PCE effects in participants' initial block, with a concurrent increase in RT and flanker effect across that block. One reason for these patterns may have been that the proportion of compatible trials differentially affected the change in RT that occurs through time (e.g., differential increases in RT leading to selectively larger flanker effects in certain conditions). As such, we next tested whether differences in RT changes through a block, as a function of the proportion compatibility in the block, explained the overall PCE. We fit a linear mixed-effects model predicting each participant's change in RT from the first 50 to the last 50 trials with the main effects of trial compatibility and block proportion compatible as well as their interaction. The model also included participant-level random effect intercepts. There were no reliable influences of a high or low proportion of compatible trials, compared to 50% compatible, on the change in flanker effect (both $|T| < 0.62$), incompatible trials (both $|T| < 0.69$), or compatible trials (both $|T| < 0.6$). Despite the clear overall differences between blocks with different proportion of compatible trials (see Fig. 1), this effect was not explained by comparisons between the flanker effects in the first 50 trials and last 50 trials. Instead, measured in this way, the change in flanker effect through time between conditions was indistinguishable. Given that the PCE must arise in response to experience with the task (i.e., statistical learning), the present analysis suggests that the changes in RT associated with the PCE likely occur quickly (e.g., within the first 50 trials; see *Finer Time Scales* section below; cf. Abrahamse et al., 2013).

Does previous experience with one block of the task affect subsequent within-block change in flanker effect?

In previous work, Abrahamse et al. (2013) observed asymmetric carryover effects in the PCE using a Stroop task. To assess whether we also observed similar carryover effects in an arrow-flanker type task, we examined whether, when compared with first blocks with the same compatibility proportion,

RT changes in subsequent blocks of trials are influenced by prior task experience. Figure 2 shows the distributions of these second-block RTs separated by participants' first blocks. We predicted that prior experience with a high-compatibility block would mean that participants would begin their second block with a relatively large flanker effect, thereby leading to a disproportionate decrease in flanker effect across the second block. The opposite pattern was predicted for blocks following the low-compatibility first blocks.

We fit a linear mixed-effects model to change in flanker effect (i.e., mean of flanker effect on last 50 trials minus mean of flanker effect on first 50 trials) with fixed effects of previous block compatibility and current block compatibility while including by-subject random intercept (see Table 2). The significant *Intercept* term in this model indicates an overall increase in flanker effect over the course of a first block of flanker trials in blocks with equal numbers of compatible and incompatible trials. The same pattern is very similar in first blocks with 20% compatible, but is nonsignificantly attenuated in first blocks with 80% compatible. Further, the reliable increase in flanker effect evident in the *Intercept* effect is significantly attenuated and fully reversed, when a second block is preceded by a first block with many incompatible trials.

To clarify these effects, we separately examined within-block changes in incompatible trials and compatible trials while controlling for the current block's compatibility proportion (see Tables 3 and 4). The comparison of interest for these changes in RT for incompatible or compatible trials was whether the change is different after certain types of blocks, as opposed to participants' first block. That is, the *Intercept* parameter indicates first-block change in RT (controlling for compatibility proportion).

While an increase in incompatible-trial RT was reliable for the first block *Intercept*, the associated change of incompatible RT in the second block was attenuated only when the first block included relatively few (50% or 20%) compatible trials.

While an increase in compatible-trial RT was reliable for the first block *intercept*, the associated change of compatible RT in the second block was significantly attenuated only when the first block included an intermediate number (50%) of compatible trials. This coefficient for this intermediate block was numerically more like the low-compatibility first block than the high-compatibility first block.

Experiment 1 discussion

Experiment 1 tested for the presence of learning giving rise to the PCE in initial and subsequent blocks of an arrow-flanker task in the absence of feedback. First, consistent with previous work, in Experiment 1 we found evidence for a PCE. The magnitude of the flanker effect was smaller in blocks with larger numbers of incompatible trials. Second, while we observed that the final magnitude of the flanker effect did emerge

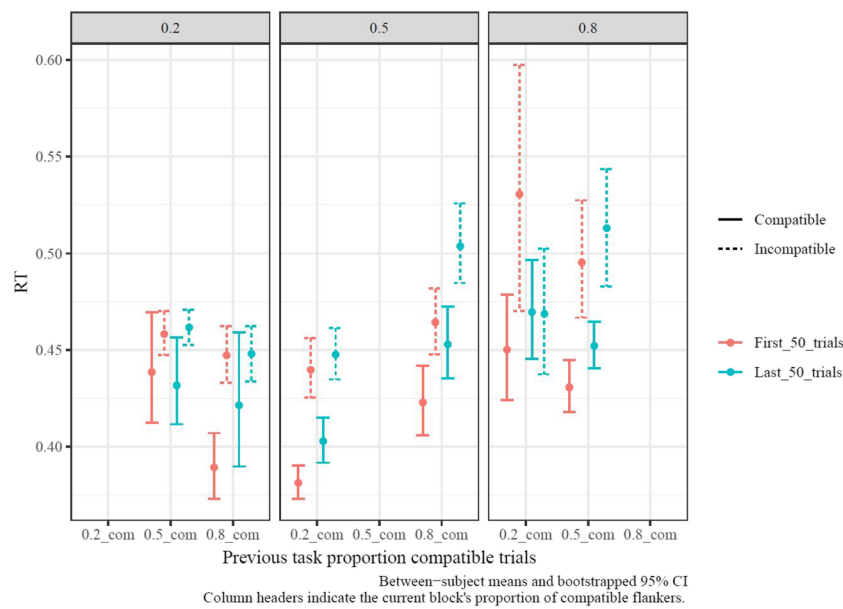


Fig. 2 Block 2 RT in the absence of feedback, separated by the first 50 and last 50 trials. As with participants’ first blocks, RT increases in several cases. When the first block had 20%-compatible trials, second block was more likely to show a pattern of converging compatible and

incompatible RT. Column headers indicate current block’s proportion compatible; *x*-axes indicate prior block’s proportion compatible. Between-subjects means and 95% CI are indicated

through time, the direction of this change through time was not consistent with a learning effect. Instead, within the first block there was a tendency toward ever slower RTs—with this effect being magnified in incompatible trials. Furthermore, we found that this drift toward slower RTs was similar across percent compatible conditions, and thus differences in a slow drift in RTs across conditions did not appear to explain the PCE. Finally, we found reliable differences in RT change in participants’ second blocks as a function of their experiences in their first blocks. Specifically, while there was an overall pattern of incompatible RTs increasing over the course of participants’ blocks, more previous experience with incompatible trials led to a subsequent attenuation of within-block RT change in the second block. This, in turn, led to the

overall within-block increases in flanker effect being significantly reversed in blocks following a 20%-compatible block. More previous experience with incompatible trials also led to an attenuation of the within-block increase in compatible-trial RT, but this effect was strongest for participants who completed 50%-compatible blocks first.

Experiment 2 results

While Experiment 1 replicated several of the core findings in the PCE literature, the overall pattern of increasing RTs was inconsistent with the changes in behavior that usually come

Table 2 Linear mixed-effects model predicting change in flanker effect for feedback-absent participants, accounting for subject-level intercepts

	Estimate (s)	SE	T	<i>p</i>
Intercept	0.022	0.011	2.104	.038
previous_20%	-0.046	0.017	-2.778	.007
previous_50%	-0.008	0.015	-0.523	.603
previous_80%	-0.021	0.017	-1.203	.234
20%_Compat	-0.003	0.015	-0.181	.858
80%_Compat	-0.021	0.015	-1.415	.163

Note. Intercept indicates change in flanker observed on first block on blocks with equal numbers of compatible and incompatible trials

Table 3 Linear mixed-effects model predicting change in incompatible-trial RT, accounting for subject-level intercepts

	Estimate (s)	SE	T	<i>p</i>
Intercept	0.054	0.016	3.467	.001
previous_20%	-0.068	0.024	-2.786	.007
previous_50%	-0.047	0.022	-2.109	.038
previous_80%	-0.024	0.025	-0.951	.346
20%_Compat	-0.007	0.021	-0.323	.749
80%_Compat	-0.006	0.021	-0.290	.774

Note. Fewer compatible trials in the first block systematically led to a smaller increase in incompatible-trial RT in the second block

Table 4 Linear mixed-effects model predicting change in compatible-trial RT, accounting for subject-level intercepts

	Estimate (s)	SE	T	<i>p</i>
Intercept	0.032	0.013	2.377	.020
previous_20%	−0.022	0.021	−1.042	.303
previous_50%	−0.039	0.019	−2.050	.044
previous_80%	−0.003	0.021	−0.153	.879
20%_Compat	−0.004	0.018	−0.233	.817
80%_Compat	0.014	0.018	0.790	.435

with experience (Newell & Rosenbloom, 1981; Schmidt, 2016; Wenke et al., 2015). That is, while participants clearly were adjusting their behavior in the task through time, these adjustments did not manifest as a decrease in RT (i.e., as might be expected from learning). More importantly for the present work, compatibility effects likewise increased. One possible reason for the failure to observe a decrease in RTs, and in the compatibility effect, through time may be a lack of feedback helping participants to modulate their behaviors. In Experiment 2, we thus implemented the same methods as in Experiment 1, with the exception that we provided participants with both response time and accuracy feedback. After each trial, participants saw a screen with their response time on the previous trial. The text was colored red if the response was incorrect or reaction times were slow (longer than 500 ms). Otherwise, the text was blue. This feedback was presented for the first 750 ms of the total 850-ms intertrial interval (see Methods). Such by-trial informative feedback is typical of PCE experiments (e.g., Gratton et al., 1992; Lehle & Hübner, 2008), and the inclusion of feedback thus better situates the current work into the broader field of PCE research.

Was the canonical PCE observed?

As with Experiment 1, we first tested for the overall expected positive interaction between trial compatibility and proportion compatible, across blocks and proportions congruent and accounting for participant-level random-effect intercepts. This predicted PCE should manifest as positive coefficients for the interaction between trial type (i.e., incompatible RT) and higher proportions compatible, possibly accompanied by negative coefficients associated with higher proportions compatible (i.e., faster compatible RT). When compared with a block with 20%-compatible trials, increasing the proportion of compatible trials reliably decreased RT on compatible trials while increasing RT on incompatible trials (see Table 5). This led to the PCE, an increase in the flanker effect in conditions with a larger proportion of compatible trials. Thus, at this broad level

Table 5 Entire-block proportion flanker effect on compatible and incompatible trials

	Estimate (s)	SE	T
Intercept	0.380	0.005	76.484
50%_compat	−0.009	0.002	−4.692
80%_compat	−0.015	0.002	−8.450
IsIncompatible	0.035	0.002	20.917
50%_compat:IsIncompatible	0.005	0.002	2.450
80%_compat:IsIncompatible	0.020	0.002	8.008

of analysis we replicated the results of Experiment 1. Unlike Experiment 1, though, the magnitude of both effects continued to increase from 50% compatible (.005 s) to 80% compatible (.020 s; see “Comparison of Experiments 1 and 2” for more explicit cross-experiment comparisons).

Does the flanker effect reliably change over the course of participants’ first 400 trials?

In Experiment 1, we observed a reliable change in both RT and the flanker effect within the first 400 trials. However, the direction of this change was such that participants got progressively slower throughout the block (disproportionately so for incompatible trials, leading to an increased compatibility effect). We surmised that this may have been due to the lack of explicit trial-by-trial feedback. We thus predicted that the introduction of feedback would, at a minimum, eliminate the drift toward slower RTs throughout the first block. We additionally expected a decrease in RT for both compatible and incompatible trials in the first block, independently of proportion compatible (as would be expected by learning). Further, disproportionate learning on incompatible trials should lead to a reduction in flanker effects from the beginning to the end of the block. We tested for these overall effects using a linear mixed-effects model predicting change in RT in participants’ first blocks with the fixed effect of trial type while controlling for participant-level intercepts.

Participants exhibited no change in flanker effect in their first block ($b = -0.001$, $T = -0.157$, $p = .875$). There was likewise no overall reliable change in either incompatible trials ($b = -0.009$, $T = -0.44$, $p = .658$) or compatible trials ($b = -0.009$, $T = -0.51$, $p = .614$). Thus, while the increases in RT we observed in Experiment 1 were no longer evident in the presence of feedback, participants in Experiment 2 did not demonstrate learning (i.e., reduction of RT; for explicit comparisons between experiments see “Comparison of Experiments 1 and 2”).

Do block compatibility proportions influence changes in response time?

One possible reason for a lack of overall change in RT in the first block may be that certain combinations of compatibility proportions are more conducive to learning than are others. Specifically, the overall PCE led us to predict that smaller proportions of compatible trials may be more associated with decreases in RT and flanker effect. This would be evident in conditions with fewer compatible trials having reliably larger decreases (i.e., more negative coefficients) in RT and flanker effect from the beginning of blocks to the end. In participants' first blocks, there was no reliable influence of high or low proportions compatible trials, compared to 50% compatible, on the change in flanker effect (both $|T| < 0.39$), incompatible trials (both $|T| < 0.98$), or compatible trials (both $|T| < 0.89$). While the numerical patterns of RT increases and decreases appeared consistent with our predictions (see Fig. 3), these between-group differences in RT change did not lead to statistically reliable differences when 80%-compatible or 20%-compatible blocks were compared to 50%-compatible blocks. As with Experiment 1, we found reliable between-condition PCE when testing all trials, but our analyses of the first and last 50 trials' RT were insensitive to changes in flanker effect giving rise to the PCE (see Finer Time Scales section for further tests of these changes).

Does previous experience with a task alter within-block change in flanker effect?

Despite the lack of evident learning when comparing the first 50 and last 50 trials of participants' first blocks, it remains possible

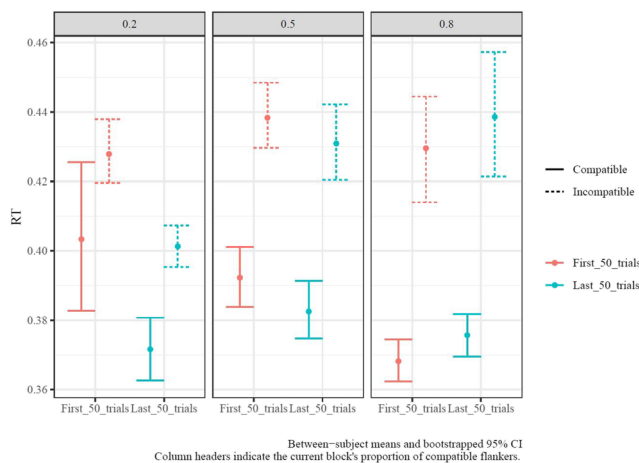


Fig. 3 Block 1 RT in the presence of feedback, separated by the first 50 and last 50 trials. Unlike in Experiment 1, there is no overall trend toward slower RT on later trials. Because changes for compatible trials were similar to changes for incompatible trials within blocks of the same proportion compatible, there was no reliable change in flanker effect. Column headers indicate current block's proportion compatible. Between-subjects means and 95% CI are indicated

that task learning may have carried over and interacted with performance on the second block. Specifically, if first-block learning-induced initial compatibility-effect changes in subsequent blocks, this may have provided exaggerated or depressed development of the PCE in the second blocks (see Fig. 4).

We tested for the effects of prior compatibility experience by predicting the change in flanker effect (i.e., mean on last 50 trials minus mean on first 50 trials) with blocks' proportion compatible and the previous proportion compatible (i.e., none/*Intercept* if predicted data were from the first block, or the first block's proportion of the predicted data were from the second block).

Changes in flanker effect were not reliable. We did not find any systematic differences in changes in flanker effect due to prior experience or due to current proportion compatible (each controlling for the other). Next, we separately tested for within-block changes in incompatible trials and compatible trials while controlling for the current block's compatibility proportion. Note the negative *Intercept* coefficients, indicating a trend toward overall improvements during the first block (see Fig. 3).

Neither compatible nor incompatible RT demonstrated change that was reliably affected by current or previous proportion compatible trials. The patterns of RT increase or decrease, although not statistically reliable, can also be observed in Fig. 4.

Experiment 2 discussion

Experiment 2 tested for learning-related changes in the PCE in two blocks of an arrow-flanker task feedback when feedback was provided. First, as with Experiment 1, in Experiment 2 we replicated the canonical PCE effect when implementing the flanker effect with feedback. Yet while the magnitude of flanker effect was different when comparing blocks with different task statistics, our measurements of change were insensitive to any within-block changes in RT (i.e., comparing the means of the first vs. last 50 trials). This stands in contrast to Experiment 1 wherein we found systematic increases in RT alongside the overall PCE (see the next section, "Comparison of Experiments 1 and 2"). Experiment 2 instead provided evidence for relatively stable response time distributions for both compatible and incompatible trials when participants were provided with feedback. The fact that the PCE was observed both in an experiment (Experiment 1) where overall RTs were drifting upward across a block and also where overall RTs were reasonably stable across a block (Experiment 2) means that it remains unclear how condition-level differences in the PCE arise through time. After direct comparisons of Experiments 1 and 2, we explore several analyses that are more sensitive to changes in RT on time scales that the previous analyses may not have been able to capture.

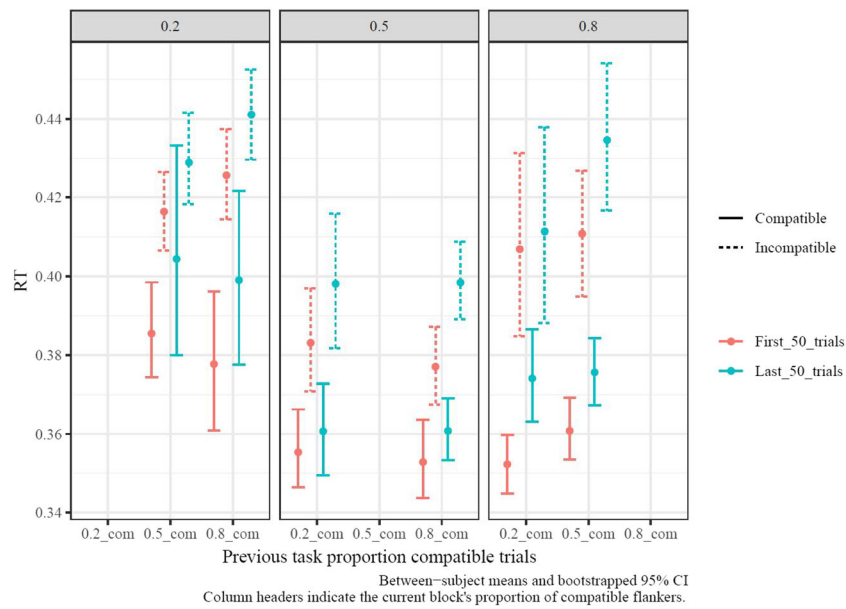


Fig. 4 Block 2 RT in the presence of feedback, separated by the first 50 and last 50 trials. There was no reliable trend of learning (i.e., decreasing RT) observed. Column headers indicate current block’s proportion

compatible; x-axes indicate prior block’s proportion compatible. Between-subjects means and 95% CI are indicated

Comparison of Experiments 1 and 2

We next included both experiments in combined models. The first, testing for PCE across all trials (i.e., combining the analyses of Tables 1 and 5), showed the PCE in the form of reliable differences between 20%-compatible flanker effects and both the 50%-compatible and 80%-compatible flanker effects (both $|T| > 3.4$). However, there was no interaction with feedback condition (both $|T| < 0.8$), indicating overall similar PCE regardless of the presence or absence of feedback.

We next specifically tested for differences between the changes in PCE in the presence or absence of feedback. As in the analyses reported in Tables 2 and 6, we test the possible feedback-related divergences in beginning-to-ending changes in the PCE (see Table 9 and Supplemental Information Fig. S3). Feedback was centered (i.e., feedback-absent = $-.5$, feedback-present = $.5$). We found no reliable effects (all

interactions with feedback $|T| < 1.25$). In contrast, in a model predicting changes in RT separated by trial type, only the main effect of feedback was reliable ($b = -0.043$, $T = -2.39$, $p = .022$; see Table S1 in the Supplementary Information). That is, changes in response times were reliably more positive by .043 s in the absence of feedback than in the presence of feedback on the baseline (first) blocks, and this effect did not reliably interact with trial compatibility or current proportion compatible. There were likewise no reliable differences between this first-block effect and any second blocks (i.e., influence of previous proportion compatible did not interact with feedback in any second-, third-, or fourth-order interactions (Tables 7 and 8)).

The differences between feedback conditions reported above appears largely to be due to increases in RT within the first blocks of participants in the feedback-absent condition (see Figs. S2–S3 in the Supplementary Information).

Table 6 Linear mixed-effects model predicting change in flanker effect for feedback-present participants, accounting for subject-level intercepts

	Estimate (s)	SE	T	p
Intercept	0.005	0.009	0.578	.566
previous_20%	-0.001	0.012	-0.114	.910
previous_50%	0.005	0.012	0.435	.666
previous_80%	0.003	0.012	0.208	.836
20%_Compat	-0.014	0.011	-1.277	.208
80%_Compat	-0.003	0.011	-0.275	.786

Note. Intercept indicates change in flanker observed on first block

Table 7 Linear mixed-effects model using prior task compatibility proportions to predict change in incompatible-trial RT, accounting for subject-level intercepts

	Estimate (s)	SE	T	p
Intercept	-0.009	0.014	-0.608	.547
previous_20%	0.017	0.019	0.913	.367
previous_50%	0.032	0.020	1.620	.111
previous_80%	0.033	0.020	1.682	.098
20%_Compat	-0.022	0.018	-1.256	.216
80%_Compat	0.014	0.018	0.808	.426

Table 8 Linear mixed-effects model using prior task compatibility proportion to predict change in compatible-trial RT, accounting for subject-level intercepts

	Estimate (s)	SE	T	<i>p</i>
Intercept	−0.015	0.013	−1.132	.262
previous_ 20%	0.020	0.018	1.126	.265
previous_ 50%	0.025	0.018	1.386	.171
previous_ 80%	0.031	0.018	1.734	.088
20%_Compat	−0.006	0.016	−0.342	.735
80%_Compat	0.017	0.016	1.042	.304

Given no learning-based account of diminishing performance with time (i.e., increased overall RT), we find it more likely that vigilance was more difficult for participants to maintain in the absence of feedback (Langner & Eickhoff, 2013). Learning effects in overall RT may have been generally absent due to the simple nature of the tasks and the prepotent responses associated with arrows. We note, however, that our investigation of possible modulations of the PCE are not necessarily contingent of specific directional changes in RTs of certain conditions or trial types (e.g., decreases in the compatibility effects could occur through relative slowing of compatible trials or through relative speeding of incompatible trials). As such, the reliable feedback-related difference in overall changes in RT is not directly informative to these questions of PCE. In contrast, there was a reliable change in compatibility effect in the absence of feedback (see Intercept term of Table 2) and a lack of reliable change in compatibility effect in the absence of feedback (see Intercept term of Table 6). This difference between conditions was not reliable, however (see Table 9).

Evaluating how the PCE emerges over the course of experimental blocks on finer time scales

In two experiments, we found robust evidence for the influence of compatibility proportion on overall response times, with less evidence for systematic within-block changes in response times. This is counterintuitive because such an effect seemingly must be a function of time. In other words, participants do not have prior knowledge of the upcoming task statistics, so their response on the very first trial of a block cannot be impacted by the distribution of trials to come. Thus, at a minimum, the effects must evolve after the very first trial. Yet, despite prior research indicating the continuation of this development over the course of hundreds of trials (Abrahamse et al., 2013), our comparisons of blocks' first 50 trials to the

last 50 trials were insensitive to the development of the PCE. Furthermore, the presence of the PCE both in Experiment 1, where within blocks we observed overall slow increases in RT, and Experiment 2, where no such changes were observed, suggests the need to look at other time scales. Given this, we conducted a set of analyses using more time-sensitive measures of the development of the PCE. We first iteratively tested the trial number on which the flanker effect differs by proportion compatible when predicting RTs. Due to the noise inherent in this analysis, we next used a parametric model of RT change (i.e., assuming monotonic change) in similar analyses. Last, we tested whether our results were accounted for by contingencies on a trial-to-trial time scale.

To examine RTs at a finer time scale, we first iteratively fit mixed-effects models predicting raw RTs on the first trial, second trial, and so on. These models included fixed effects of current block proportion compatible, current trial proportion compatible, and the interaction between these two. The models also controlled for previous block proportion compatible and participant-level intercepts. Low-compatibility blocks had flanker effects that diverged from the high-compatibility block quickly, with the flanker effect appearing different by block type (i.e., interaction between block compatibility and trial compatibility; $T > 2$) by trial 13 in Experiment 1 and trial 22 in Experiment 2 (see Fig. 5). This is consistent with the idea that these effects appear quite quickly in a block. Yet we note that these estimates also appear to be heavily influenced by sampling noise. The reported test would likely have a false positive rate of .05, leading to approximately 1 out of 20 tests appearing reliable by chance. We note here that we did not correct for multiple comparisons because this set of models was not intended as a statistical test per se, but rather intended to demonstrate the trials on which conditions would be interpreted as diverging if only that data were taken into account (and by our best estimates, this divergence occurs very quickly in the experiment). If standard multiple comparison corrections were applied, no trial-wise comparison would remain significant (i.e., there is a high likelihood of false positives across this set of models).

To mitigate noise in this analysis, we conducted the same iterative tests on nonlinear model predictions of the flanker effect. We first fit a maximum likelihood exponential learning model to each participant's flanker effect for each block using the R package TEfits (Cochrane, 2020), producing by-trial estimates of the flanker effect for each participant and block (see [Supplementary Information](#) for model details). As with the analyses above treating each trial as independent, trial-level estimates of participants' performance were noisy, but averaged model estimates across participants demonstrate a smooth continuous development of the PCE over time. The trend in the PCE, however, appears to be an increase over time

Table 9 Linear mixed-effects model using prior task compatibility proportion to predict change in flanker effects on RT, accounting for subject-level intercepts

	Estimate (s)	SE	T	<i>p</i>
Intercept	0.006	0.008	0.804	.422
prevTask0.2_com	−0.007	0.014	−0.484	.629
prevTask0.5_com	−0.008	0.014	−0.530	.597
prevTask0.8_com	0.004	0.014	0.289	.772
propConFactor0.2	0.002	0.012	0.184	.853
propConFactor0.8	0.002	0.012	0.151	.879
feedback	−0.013	0.016	−0.820	.412
prevTask0.5_com:propConFactor0.2	0.003	0.02	0.134	.895
prevTask0.8_com:propConFactor0.2	−0.024	0.021	−1.130	.26
prevTask0.2_com:propConFactor0.8	−0.038	0.021	−1.797	.074
prevTask0.2_com:feedback	0.033	0.028	1.206	.229
prevTask0.5_com:feedback	0.026	0.029	0.908	.365
prevTask0.8_com:feedback	0.013	0.028	0.456	.649
propConFactor0.2:feedback	−0.017	0.023	−0.704	.482
propConFactor0.8:feedback	0.005	0.024	0.209	.834
prevTask0.5_com:propConFactor0.2:feedback	−0.016	0.04	−0.400	.689
prevTask0.8_com:propConFactor0.2:feedback	0.021	0.043	0.491	.624
prevTask0.2_com:propConFactor0.8:feedback	0.036	0.042	0.852	.395

in the presence of feedback and a slight decrease in the absence of feedback (after an initial increase). This indicates that feedback may enhance, or at least encourage the maintenance of, learned patterns of attentional allocation giving rise to the PCE. These differences do not appear to vary in their time course of onset.

We used these model estimates in an analysis using the same iterative by-trial models described above, and again we found differential flanker effects emerging very quickly in both experiments (in this analysis, before Trial #4; see

Fig. 5). Thus, in analysis of raw data as well as data fit with learning models, these time scales are clearly shorter than what is necessary to effectively learn from task compatibility proportion (i.e., there is little possibility of learning that a block has 80%-compatible trials in only four total trials). Further, when testing for differences between the parameters themselves of the nonlinear models (i.e., starting flanker effect, rate of change, or asymptotic flanker effect), no reliable effects were evident of current or previous proportions compatible. This indicates that, like aggregating over the first and

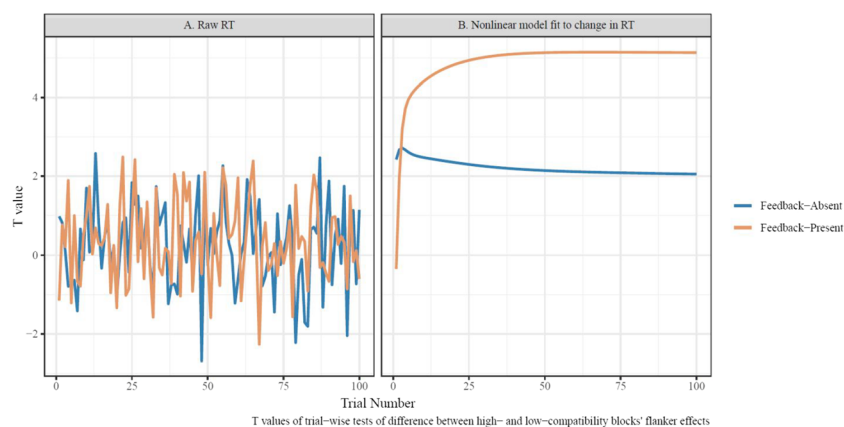


Fig. 5 Time course of change in flanker effect between low-compatibility and high-compatibility conditions over the first 100 trials, as tested by linear mixed-effects models fit to each of the first 100 trial numbers. **a** When testing differences in raw RT, differences are noisy between high-

compatibility and low-compatibility blocks. **b** When testing differences in RT fit by nonlinear regression, the rapid differences between conditions are clearer

last 50 trials, by-trial learning models reported here are unable to fully capture the time scale of change associated with the PCE.

As noted in the Introduction, another possible root of the block-wise PCE is not necessarily based upon learning per se, but simple trial-to-trial contingencies in behavior. For example, flanker effects may be smaller after incompatible trials (Braem et al., 2019; Davelaar & Stevens, 2009; Gratton et al., 1992, cf. Duthoo, Abrahamse, Braem, Boehler, & Notebaert, 2014). The PCE would then emerge because there are simply more incompatible trials in some blocks. Using linear mixed-effects models to test the interaction between current and previous-trial compatibility, we found that this effect does exist overall in both feedback-absent ($b = -0.011$, $T = -3.96$) and feedback-present conditions ($b = -0.009$, $T = -5.21$).

Interestingly though, the links between these rapid trial-wise changes in behavior and longer within-block or cross-block effects remained unclear. For instance, in mixed-effects models for each experiment's first block predicting RT with previous trial's compatibility, current trial's compatibility, overall proportion compatible, and second-order interactions between current trial compatibility and the other predictors, the interactions between block proportion compatible and current trial compatibility generally remained reliable even when controlling for the other effects (with 50% compatible as a reference; without-feedback 20% $T = -1.25$, 80% $T = 2.56$; with-feedback 20% $T = -3.2$, 80% $T = 3.1$). These results show that decreased (20%-compatible blocks) or increased (80%-compatible blocks) flanker effects, when compared to the 50%-compatible block, are not fully accounted for by trial-to-trial contingencies. Thus, while RTs are reliably influenced by the previous trial's compatibility, the PCE was still evident in our data when controlling for this short-term effect.

General discussion

Here, we tested the time course of adaptation of attentional control to task statistics. Following Eriksen (C. W. Eriksen, 1995), our findings broadly support a view of visual selective attention as rapidly contextually modulated via scaling inhibition. First, we replicated the canonical proportional compatibility effect (PCE) across two experiments using an arrow-flanker task, the first without feedback and the second with feedback. In each experiment, RTs on incompatible trials tended to be slower in blocks with fewer incompatible trials. There was a corresponding effect on compatible trials, where-in RTs were longer in blocks with fewer compatible trials.

However, some patterns diverged when examining the temporal dynamics of performance in these two experiments. In Experiment 1, which did not include explicit performance feedback, participants' RTs reliably increased over the course

of the first block of trials. This increase runs contrary both to the prediction of a low-level learning model (i.e., where RTs should generally decrease with experience with a task), as well as one of the primary justifications for omitting feedback in the broader attention and cognition literature (i.e., that omitting feedback would produce more stable behavior). More relevant to compatibility effects and the PCE, theories of learning and adaptive control that address the PCE each predict optimization of performance in the form of reduced response times (Braem et al., 2019). In contrast to these perspectives, our results showed that participants' first feedback-absent blocks were associated with significant increases in incompatible-trial RT and in compatible-trial RT, leading to a significant increase in flanker effect over time. When feedback was present, no changes in RT or flanker effect were reliably evident when comparing the first 50 trials to the last 50. In direct comparisons of the two experiments, however, only the overall RT changes were reliably influenced by the presence of feedback, with the changes of flanker effect lacking any reliable difference between the two experiments. The drift toward slower RTs in the absence of feedback could be caused by a number of mechanisms, including issues related to sustained attention, fatigue, or an attempt to find a stable compromise between speed and accuracy—as any of these could be countered by trial-by-trial feedback, as in Experiment 2. While the extent to which such drifts toward slower RTs are seen in the literature (where typical analysis approaches average over all trials within a block and thus would not detect such a drift) may be of interest, for the perspective of this current work, it is more relevant to consider the fact that the key measures of interest in the task (PCE, flanker compatibility effect) are observed in two situations where the global pattern of results are quite different from one another (i.e., Experiment 1, where global RTs become slower through time; Experiment 2, where global RTs are reasonably stable through time).

Further, in Experiment 1, patterns of RT change in sequential blocks showed that experience with more incompatible trials in an initial block led to a reliable attenuation of flanker-effect increase in subsequent blocks. That is, when participants completed an initial block with only 20%-compatible trials, their second block's flanker effect was likely to decrease over time rather than increase. This effect was largely due to RTs in incompatible trials; within-block increases in incompatible-trial RTs were significantly attenuated when blocks were preceded by 20% or 50% compatible blocks. The magnitude of this attenuation was approximately linear across 20%, 50%, and 80% compatible blocks. Compatible-trial attenuation of RT increase follows a different and nonmonotonic pattern in which only the effect of being preceded by a 50%-compatible block is significant.

In all, the results of Experiment 1 indicate global (i.e., above and beyond local statistics) learning from experience

with incompatible trials such that second blocks' incompatible trials are improved with, but not without, majority-incompatible initial blocks. This learning cannot be wholly explained by carryover from Block 1 to Block 2—if second blocks simply demonstrated initially lowered RT for incompatible trials due to previous experience with many incompatible trials, RT would not be expected to decrease even further within these second blocks. In other words, second-block decreases in incompatible RT seem to attenuate the PCE rather than being a result of the PCE. However, each of these interpretations is clouded by the overall increases in compatibility effects observed in Experiment 1. In RT measures, learning is typically considered to be manifested in the form of decreased RT. There is no necessity in this relationship, though, and it is clearly possible that any task may be associated with learning, which does not influence the measured behavior (i.e., measurements may be inadequate indicators of internal states). By testing sequential learning, we were able to observe learning effects that were not evident in within-block measures.

Despite replicating the canonical PCE effect, the patterns of change in Experiment 1 were unexpected, given standard theories of learning (i.e., increasing overall RTs, with a disproportionate increase in incompatible trials resulting in increasing flanker effects). As such, despite observing results that would be considered consistent with the PCE, it may be possible that our pattern of results in Experiment 1 arose from processes of change that are not typical of PCE results. One reason for the differences may have been that Experiment 1, unlike most PCE research, involved the use of flanker tasks without feedback. PCE is most often studied in Simon-like or Stroop-like tasks (e.g., Hutchison, 2011; Spinelli et al., 2019; Wühr, Duthoo, & Notebaert, 2015) and/or using feedback (e.g., Lehle & Hübner, 2008; Wendt, Kluwe, & Vietze, 2008). It is possible that the error signals regarding incorrect or slow trials, whether explicit or self-monitoring, that participants receive in a no-feedback flanker task are weaker than the more commonly studied paradigms and are therefore unable to drive learning in the form of overall decreases in RT. In this context, *error* would be any noncompliance with the instructions to complete the task quickly and accurately. The mechanism by which the error signal would act (e.g., narrowed scope of attention, increased engagement) is not specifically of interest here. Instead, we simply wanted to align Experiment 2 with a set of theories that assumes that participants receive strong feedback signals. To investigate the possibility that increased error signal would lead to more canonical patterns of RT change over time, we tested the sensitivity of the observed learning to the inclusion of by-trial feedback.

In Experiment 2 we found that, unlike in Experiment 1, RT on compatible trials as well as RT on incompatible trials decreased over the course of each block, although neither of these effects was statistically reliable. Compatible RT

decreased more than incompatible RT, leading to a nonreliable increase in flanker effect. Apart from the overall PCE, no statistically reliable effects were observed in any of the Experiment 2 data. This lack of reliable effects is remarkable, given the robustness of the PCE; proportion-flanker effects must necessarily develop over time with accumulated experience with task statistics, yet the measures used here are insensitive to these changes. We next used by-trial nonlinear regressions of the flanker effect fit to each block of each participant. In comparing the coefficients and predictions from these models, we still found no reliable condition differences in changes over the course of blocks. This indicates that the evident change in RT distributions must have occurred rapidly enough that even by-trial learning models are unable to effectively capture the change. The PCE remains reliable even when controlling for adjacent-trial effects, however, indicating that the PCE we observed has a source above and beyond single-trial fluctuations. Unlike previous reports demonstrating carryover of attentional changes between adjacent blocks of a response competition task (Abrahamse et al., 2013), we found inconsistent evidence for persistent changes on the time scale of blocks (i.e., hundreds of trials). Instead, the mechanisms of attentional modulation were likely to be occurring on short time scales that did not lead to consistent cross-block influences.

Many of the results reported here have been null effects which preclude strong inferences. Statistical evidence for the null hypotheses tended to be strong (i.e., Bayes factors, approximated using Bayesian information criterion (BIC) comparisons, with over 10 times as much evidence for the null hypothesis than the alternative hypothesis; Wagenmakers, 2007). Nonetheless, a lack of a clear interpretation of null effects remains due to underspecified causes for a lack of difference between beginning and ending PCE (e.g., adjacent-trial effects do not explain our results).

While the bias in task statistics (i.e., that there are more compatible trials or more incompatible trials in a given block) can necessarily only emerge through time, our by-trial learning models indicated that the PCE appears to emerge very early. While such a result obviously does not rule out longer-term phenomena, it does provide support for theories that have suggested a driving role for more short-range effects (e.g., at the level of trial-to-trial effects). Indeed, our results do suggest that such short-range effects cannot explain the entirety of the PCE (i.e., that there is variance explained by task statistics that is uncaptured by such short-range effects). To better differentiate possible processes (e.g., longer-range learning of statistics versus trial-to-trial effects) it may be valuable for future work to manipulate either the difficulty of learning task statistics (i.e., to slow the long-range processes, thus making them more distinct from short-range processes) or to implement manipulations that should reduce the short-range (trial-to-trial) effects. One method for inhibiting the rate

of statistical learning may be to reduce the contrast between blocks' biases (e.g., contrasting 65% congruent to 35% congruent rather than 80% and 20%). While this method would likely slow down learning, differences between such a study and the current results would be confounded by the possibility of reduced magnitudes of PCE due to smaller contrasts. Another method may be to insert pauses or distracting tasks at occasional intervals (e.g., every 20 trials) to make learning the long-range statistics more difficult. In terms of inhibiting short-range (e.g., trial-to-trial) effects, possibilities would be to include distracting stimuli or dual task demands fully interleaved with the primary response competition task (i.e., after every trial; although it is not clear whether it is possible to disrupt the short-range interactions without, at the same time, disrupting the ability to learn the long-range statistics). We note that the feedback introduced in Experiment 2 could have had such an effect (i.e., of introducing a visual stimulus not drawn from the task-relevant stimulus set), although we did not find that visual feedback served to inhibit the magnitude of the PCE nor its rate of onset. Other stronger manipulations, either passive (e.g., masking stimuli with an enforced intertrial interval) or active (e.g., an auditory discrimination task) may be more effective at slowing altering the magnitude or range of PCE onset.

Here, we have taken an approach to the study of attention that is quite different than most PCE studies. In two experiments, we replicated the PCE effect when averaged across all trials as well as within-trial pairs (i.e., smaller flanker effects on trials following incompatible trials). However, given the overall goal of integrating the PCE effect into a learning framework that would be applicable to broader fields of attention and cognition, our results are mixed. As with PCE results more generally, the systematic changes in flanker effect we observed should act as a contextualizing caution to researchers putting a large stake on single measures of response competition. Our reported divergence between feedback-present and feedback-absent experiments should provide a basis for future research to consider the combination of feedback and compatibility best suited to answer questions of response competition, individual differences, learning, or fatigue. In addition, the possible influence of sequential task demands cannot be dismissed even in this fairly simple task.

Conclusion

Our work corroborates the broader PCE literature in providing unequivocal evidence for interacting bottom-up (i.e., stimulus-driven) and top-down (e.g., learning, attentional modulation) processes in response competition. Despite our goals of identifying sources of variation in the magnitude and time scale of PCE modulations, our measures were insensitive to

the time course of learning. In contrast with previous work positing the development of the PCE over hundreds of trials (Abrahamse et al., 2013), we found the presence of the PCE very early when using by-trial modeling of response times. This indirectly supports a time scale of modulation as small as single trials. Nonetheless, future work would benefit from methodological innovations facilitating an identification of the interacting time scales giving rise to sequential PCE effects.

Open practices statement The data and materials for all experiments are available at 10.17605/OSF.IO/KMBZ8, and none of the experiments was preregistered.

Funding This work was supported in part by Office of Naval Research Grant ONR-N000141712049. This funding source had no direct involvement in study design, data collection, analysis, manuscript preparation, or any other direct involvement in this research.

References

- Abrahamse, E., Braem, S., Notebaert, W., & Verguts, T. (2016). Grounding cognitive control in associative learning. *Psychological Bulletin*, 142(7), 693–728. <https://doi.org/10.1037/bul0000047>
- Abrahamse, E., Duthoo, W., Notebaert, W., & Risko, E. F. (2013). Attention modulation by proportion congruency: The asymmetrical list shifting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(5), 1552–1562. <https://doi.org/10.1037/a0032426>
- Bailey, K., Amlung, M. T., Morris, D. H., Price, M. H., Von Gunten, C., McCarthy, D. M., & Bartholow, B. D. (2016). Separate and joint effects of alcohol and caffeine on conflict monitoring and adaptation. *Psychopharmacology*, 233(7), 1245–1255. <https://doi.org/10.1007/s00213-016-4208-y>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>
- Braem, S., Bugg, J. M., Schmidt, J. R., Crump, M. J. C., Weissman, D. H., Notebaert, W., & Egner, T. (2019). Measuring Adaptive Control in Conflict Tasks. *Trends in Cognitive Sciences*, 23(9), 769–783. <https://doi.org/10.1016/j.tics.2019.07.002>
- Bugg, J. M., & Crump, M. J. C. (2012). In support of a distinction between voluntary and stimulus-driven control: A review of the literature on proportion congruent effects. *Frontiers in Psychology*, 3, 367. <https://doi.org/10.3389/fpsyg.2012.00367>
- Cochrane, A. (2020). TEfits: Nonlinear regression for time-evolving indices. *Journal of Open Source Software*, 5(52), 2535. <https://doi.org/10.21105/joss.02535>
- Cochrane, A., Simmering, V. R., & Green, C. S. (2019). Fluid intelligence is related to capacity in memory as well as attention: Evidence from middle childhood and adulthood. *PLOS ONE*, 14(8), e0221353. <https://doi.org/10.1371/journal.pone.0221353>
- Davelaar, E. J., & Stevens, J. (2009). Sequential dependencies in the Eriksen flanker task: A direct comparison of two competing accounts. *Psychonomic Bulletin & Review*, 16(1), 121–126. <https://doi.org/10.3758/PBR.16.1.121>
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Duthoo, W., Abrahamse, E. L., Braem, S., Boehler, C. N., & Notebaert, W. (2014). The congruency sequence effect 3.0: A critical test of

- conflict adaptation. *PLOS ONE*, 9(10), e110462. <https://doi.org/10.1371/journal.pone.0110462>
- 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.
- Eriksen, C. W. (1995). The flankers task and response competition: A useful tool for investigating a variety of cognitive problems. *Visual Cognition*, 2, 101–118.
- 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. <https://doi.org/10.1162/089892902317361886>
- Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, 121(4), 480–506.
- Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models—The R package pbkrtest. *Journal of Statistical Software*, 59(9). <https://doi.org/10.18637/jss.v059.i09>
- Hutchison, K. A. (2011). The interactive effects of listwide control, item-based control, and working memory capacity on Stroop performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(4), 851–860. <https://doi.org/10.1037/a0023437>
- 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. <https://doi.org/10.1016/j.cub.2017.01.046>
- Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin*, 139(4), 870–900. <https://doi.org/10.1037/a0030694>
- Lehle, C., & Hübner, R. (2008). On-the-fly adaptation of selectivity in the flanker task. *Psychonomic Bulletin & Review*, 15(4), 814–818.
- Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & Cognition*, 7(3), 166–174. <https://doi.org/10.3758/BF03197535>
- Miyake, A. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49–100. <https://doi.org/10.1006/cogp.1999.0734>
- Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–51). Hillsdale, NJ: Erlbaum.
- Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, 35(1), 73–89. <https://doi.org/10.1146/annurev-neuro-062111-150525>
- Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, 13, 25–42.
- Rueda, M. R., Rothbart, M. K., McCandliss, B. D., Saccomanno, L., & Posner, M. I. (2005). Training, maturation, and genetic influences on the development of executive attention. *Proceedings of the National Academy of Sciences of the United States of America*, 102(41), 14931–14936. <https://doi.org/10.1073/pnas.0506897102>
- Schmidt, J. R. (2016). Proportion congruency and practice: A contingency learning account of asymmetric list shifting effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(9), 1496–1505. <https://doi.org/10.1037/xlm0000254>
- Schmidt, J. R., & Weissman, D. H. (2014). Congruency sequence effects without feature integration or contingency learning confounds. *PLOS ONE*, 9(7), e102337. <https://doi.org/10.1371/journal.pone.0102337>
- Seitz, A., & Dinse, H. R. (2007). A common framework for perceptual learning. *Current Opinion in Neurobiology*, 17(2), 148–153. <https://doi.org/10.1016/j.conb.2007.02.004>
- Sidarus, N., Palminteri, S., & Chabon, V. (2019). Cost-benefit trade-offs in decision-making and learning. *PLoS Computational Biology*, 15(9), e1007326. <https://doi.org/10.1371/journal.pcbi.1007326>
- Spinelli, G., Perry, J. R., & Lupker, S. J. (2019). Adaptation to conflict frequency without contingency and temporal learning: Evidence from the picture–word interference task. *Journal of Experimental Psychology: Human Perception and Performance*, 45(8), 995–1014. <https://doi.org/10.1037/xhp0000656>
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14(5), 779–804. <https://doi.org/10.3758/BF03194105>
- Wendt, M., Kluwe, R. H., & Vietze, I. (2008). Location-specific versus hemisphere-specific adaptation of processing selectivity. *Psychonomic Bulletin & Review*, 15(1), 135–140. <https://doi.org/10.3758/PBR.15.1.135>
- Wenke, D., De Houwer, J., De Winne, J., & Liefoghe, B. (2015). Learning through instructions vs. learning through practice: Flanker congruency effects from instructed and applied S–R mappings. *Psychological Research*, 79(6), 899–912. <https://doi.org/10.1007/s00426-014-0621-1>
- 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. <https://doi.org/10.1093/cercor/bhq101>
- Wühr, P., Duthoo, W., & Notebaert, W. (2015). Generalizing attentional control across dimensions and tasks: Evidence from transfer of proportion-congruent effects. *Quarterly Journal of Experimental Psychology*, 68(4), 779–801. <https://doi.org/10.1080/17470218.2014.966729>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.