Journal of Experimental Psychology: General

ISSN: 0096-3445

2024, Vol. 153, No. 6, 1644–1670
https://doi.org/10.1037/xge0001585

Tango of Control: The Interplay Between Proactive and Reactive Control

Giada Viviani[1, 2], Antonino Visalli[3], Maria Montefinese[4], Antonino Vallesi[1, 2], and Ettore Ambrosini[1, 2, 5]
[1] Department of Neuroscience, University of Padova
[2] Padova Neuroscience Center, University of Padova
[3] IRCCS San Camillo Hospital, Venice, Italy
[4] Department of Developmental and Social Psychology, University of Padova
[5] Department of General Psychology, University of Padova

Cognitive control has been theorized operating through two distinct mechanisms, proactive and reactive control, as posited by the dual mechanism of control model. Despite its potential to explain cognitive control variability, the supporting evidence for this model remains inconclusive. Prior studies frequently employed the Stroop task to assess this model, manipulating the proportion congruency (PC) at the list-wide and/or item-specific levels to target proactive and reactive control, respectively. However, these manipulations have been questioned as they may invoke low-level associative learning instead of control-driven mechanisms. Although solutions have been proposed to address these concerns, they still have limitations and impracticalities. In pursuit of a clearer understanding of this issue, we manipulated proactive and reactive control simultaneously to more directly investigate their separability. We conducted two experiments using a peripheral and a perifoveal spatial Stroop task version, respectively, and we adopted state-of-the-art methodologies, leveraging trial-level multilevel modeling analytical approaches, to effectively estimate the Stroop effect and its control-related modulations while controlling for confounding factors. Notably, we manipulated both list-wide and item-specific PCs at the trial level, allowing for a fine-grained analysis. Our results provide compelling evidence for the existence of a list-wide, PC-dependent proactive control mechanism, influencing Stroop performance independently of reactive control and confounding factors. Additionally, an item-specific PC-dependent reactive control effect was found to influence Stroop performance only in interaction with proactive control. These findings contribute to a better understanding of the interplay between proactive and reactive control mechanisms, shedding light on the intricate nature of cognitive control.

Public Significance Statement
Our study investigated how our brain manages selecting relevant information in face of distracting or conflicting stimuli according to the dual mechanism of control model, which suggests that we use two distinct strategies, proactive and reactive control. We found that proactive control is dominant when handling cognitive conflict. This understanding may help develop strategies for better decision making and focus in daily life.

Keywords: dual mechanism of control model, proactive and reactive control, spatial Stroop task, trial-level list-wide and item-specific proportion congruency manipulation, contingency

This article was published Online First April 25, 2024.
Aysecan Boduroglu served as action editor.
Ettore Ambrosini https://orcid.org/0000-0002-1439-5491
This study was in part supported by "Progetto giovani ricercatori" Grants from the Italian Ministry of Health (Project Code: GR-2018-12367927–FINAGE to Antonino Vallesi; Project Code: GR-2019-12371166 to Ettore Ambrosini), by the Investment line 1.2 "Funding projects presented by young researchers" (CHILDCONTROL) from the European Union -NextGenerationEU to Maria Montefinese, and by the "Department of Excellence 2018–2022" initiative of the Italian Ministry of University and Research, awarded to the Department of Neuroscience, University of Padova. All materials, data, code, and supplemental materials can be found at https://osf.io/qmu7g. A preprint version of this article can be found at https://osf.io/2s8gm.

Giada Viviani served as lead for formal analysis, investigation, writing–original draft, and writing–review and editing, contributed equally to data curation, methodology, and validation, and served in a supporting role for software and visualization. Antonino Visalli served as lead for methodology and software and served in a supporting role for data curation, formal analysis, investigation, and validation. Maria Montefinese served in a supporting role for investigation. Antonino Vallesi served as lead for resources and served in a supporting role for supervision. Ettore Ambrosini served as lead for conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, supervision, validation, visualization, and writing–review and editing and contributed equally to resources and writing–original draft. Giada Viviani and Antonino Visalli contributed equally to conceptualization. Maria Montefinese, Antonino Vallesi, and Ettore Ambrosini contributed equally to funding acquisition. Antonino Visalli, Maria Montefinese, and Antonino Vallesi contributed equally to writing–review and editing.

Correspondence concerning this article should be addressed to Ettore Ambrosini, Department of Neuroscience, University of Padova, via Belzoni, 160, 35121 Padova, Italy. Email: ettore.ambrosini@unipd.it

Cognitive control is one of the abilities that mostly characterizes human beings, as it is fundamental to goal-directed behavior (Chiew & Braver, 2017; Cohen, 2017). Cognitive control is an umbrella term that refers to a family of processes required to adaptively regulate, coordinate, and sequence our thoughts and action plans according to the context and internal goals (Braver, 2012; Chiew & Braver, 2017). Indeed, it comes into play when simple automatic behavior is not enough and more flexible or complex behavior, guided by internal states or intentions, is needed (Miller & Cohen, 2001). Cognitive control works by maintaining current goal representations to bias cognitive processes in favor of goal-directed stimuli and actions and by updating such representations when goals and context change (Botvinick et al., 2001; Chiew & Braver, 2017; Cohen, 2017; Diamond, 2013; Miller & Cohen, 2001). A central component of cognitive control is interference resolution, which is the ability to select weaker but task-relevant information when it is in competition with a stronger and more habitual, but task-irrelevant one (Miller & Cohen, 2001; Nee et al., 2007; Tafuro et al., 2019). The need to select task-relevant information among conflicting one is pervasive in everyday life because we are always surrounded by a great amount of sensory stimuli and possible actions, but only some of them are appropriate at any given moment (Gratton et al., 2018; Jiang et al., 2014; Nee et al., 2007).

In the last decades, the mechanisms that underlie and adaptively regulate cognitive control have been intensively investigated. However, an exhaustive understanding of this fundamental process is still lacking. Therefore, in the present study, we directly investigated one of the most influential cognitive control accounts, the dual mechanism of control model (DMC; Braver, 2012; Braver et al., 2007), using the Stroop task. Our aim was to assess whether the two control modes postulated by the DMC model, namely proactive and reactive control, are separable. Indeed, due to some methodological and theoretical issues, it still remains unclear whether they can be distinguished at the behavioral level. To achieve this, we devised a novel methodological approach that incorporates cutting-edge trial-level multilevel modeling techniques, ensuring accurate and reliable estimations of the Stroop effect, while finely manipulating the proxy variables for proactive and reactive control at the trial level. This fine-grained manipulation is crucial for gaining insights into the simultaneous presence and interplay of proactive and reactive control mechanisms while effectively controlling for potential confounding effects arising from low-level processes, such as contingency.

Given the vastness of the topic and the extensive literature on it, in what follows we first describe the classic manipulations commonly employed to engage proactive and reactive control, along with their respective limitations. Subsequently, we will discuss how researchers typically address such flaws and highlight the associated costs of proposed solutions. Additionally, we will highlight why the commonly used approaches do not allow truly investigating the separability of the two mechanisms. We will thus propose an alternative approach to solve these issues.

The Stroop Task: Proportion Congruency (PC) Manipulations and Their Limitations

In the laboratory, cognitive control can be investigated using interference tasks (Bugg & Crump, 2012; Gratton et al., 2018; Jiang et al., 2014). The Stroop task is among the most widely used interference tasks (Stroop, 1935). In its original version, known as the color-word Stroop task, words denoting a color are presented in either the same or a different ink color and participants are instructed to name the ink color regardless of the word meaning. Typically, the so-called Stroop effect is observed, that is, participants' performance is worse in incongruent trials—when the meaning and the ink color of the word do not match—as compared to congruent trials—when they do match (MacLeod, 1991; Stroop, 1935). Therefore, in this task, cognitive control is engaged to overcome interference at the task level (i.e., reading vs. color naming), as well as interference occurring at the stimulus and response levels in incongruent trials (Gonthier et al., 2016; Viviani et al., 2023).

Although the Stroop effect is universally observed, its magnitude varies as a function of age and clinical condition, or it can be deliberately modulated through specific experimental manipulations (e.g., Bugg & Crump, 2012). Of interest to the current work is that cognitive control demand and the consequent magnitude of the Stroop effect can be modulated by manipulating the PC, that is, the relative frequency/likelihood of congruent trials within the task blocks (Gonthier et al., 2016; Logan & Zbrodoff, 1979). The basic assumption is that information about the PC is used to adjust the cognitive control level and, as the size of the Stroop effect inversely reflects the success of interference resolution, it is postulated that when such effect is relatively small, a greater extent of cognitive control has been recruited (e.g., Braem et al., 2019; Lindsay & Jacoby, 1994). More in detail, in blocks with a low proportion of congruent trials (low PC), the high probability of encountering incongruent trials and experiencing interference increases cognitive control demands and this yields significantly smaller Stroop effects. In contrast, in blocks with a high proportion of congruent trials (high PC), due to a relatively lower frequency of incongruent trials, cognitive control is laxer, and the Stroop effect gets larger (e.g., Lindsay & Jacoby, 1994).

The high flexibility of the PC manipulation makes it suitable for our purpose of differentiating the distinct cognitive control mechanisms (Bugg, 2017; Bugg & Crump, 2012), namely, those postulated by the DMC model (Braver, 2012; Braver et al., 2007). As introduced above, the DMC model explains the intrinsic variability of cognitive control in terms of different temporal dynamics, postulating that there are two qualitatively distinct cognitive control modes: proactive and reactive control. The former mode operates actively by maintaining task goals and anticipatorily biasing cognitive processes in a goal-driven manner. Thus, proactive control acts as a preparatory mechanism, engaged in a sustained fashion even before cognitively demanding events, like conflicts, are encountered. When proactive control is exerted, interference is reduced because top-down attentional biases favor the processing of the task-relevant information. By contrast, reactive control is mobilized transiently only as needed on a "just-in-time" basis. As such, reactive control relies upon a "late correction" mechanism reflecting the bottom-up reactivation of task goals to resolve interference. Empirical evidence supports the DMC model by emphasizing the necessity for multiple control mechanisms to comprehensively account for different experimental effects and their differences in terms of time scales (Torres-Quesada et al., 2013).

Previous works using the Stroop task have shown that these two cognitive control modes can be distinguished by manipulating the PC at the list-wide and item-specific levels (Bugg, 2012; Bugg &

Crump, 2012).[1] Indeed, although these PC manipulations produce a similar overall pattern, they favor the use of a cognitive control mode over the other, as the logic behind them is different (Bugg, 2017). The list-wide PC manipulation is used to stress the adoption of a proactive control mechanism to resolve Stroop interference. It implies varying the PC within experimental blocks, that is, blocks with high list-wide PC, in which the proportion of congruent trials is higher (e.g., 75% congruent and 25% incongruent), are compared to blocks with low list-wide PC, wherein the ratio is reversed (e.g., 25% congruent and 75% incongruent). Typically, this manipulation yields the so-called list-wide PC effect, characterized by smaller Stroop effects in blocks with a low list-wide PC as compared to blocks with a high list-wide PC (e.g., Bugg & Crump, 2012; Lindsay & Jacoby, 1994; Logan & Zbrodoff, 1979). Such an effect would be yielded by a goal-driven modulation of control, which is possible because, after experiencing a number of trials within a block, participants learn the global likelihood of conflict and develop expectancies about the upcoming trials. Low list-wide PC leads to the implementation of early preparatory strategies operating even before stimulus onset, which entail imposing an attentional bias toward the task-relevant dimension and/or away from the task-irrelevant one. By contrast, when list-wide PC is high, the more prepotent task-irrelevant dimension is processed preferentially (Braver et al., 2007; Bugg, 2017; Bugg & Chanani, 2011; Bugg & Crump, 2012; Lindsay & Jacoby, 1994; Logan & Zbrodoff, 1979). The highest level of proactive control is observed in low list-wide PC blocks, yielding not only shorter response times (RTs) on incongruent trials, but also a congruency cost, namely a slowing on congruent trials as compared to congruent trials in high list-wide PC blocks, since the attentional biases away from the task-irrelevant dimension, which are imposed anticipatorily and globally, reduce the facilitation on congruent trials (Gonthier et al., 2016).

In contrast, when the item-specific PC is manipulated, reactive control is dominant. As the name suggests, it is implemented by assigning different PCs to specific sets of items (Jacoby et al., 2003). Essentially, this manipulation targets one feature of an item, which is commonly the task-irrelevant dimension (i.e., the word dimension in the color-word Stroop task). Such stimulus feature signals a specific PC and two conditions can be distinguished within the same block: items with a high item-specific PC (e.g., 75% probability for the word red to appear in red ink) and item with a low item-specific PC (e.g., 25% probability for the word blue to occur in blue ink). Therefore, low item-specific PC items, by signaling a high level of expected interference, are assumed to produce the highest level of reactive control (e.g., Bugg, 2012, 2017; Bugg & Hutchison, 2013) triggered by a fast "stimulus-attention association" (Tafuro et al., 2020). Using this manipulation, previous studies found an item-specific PC effect, that is, smaller Stroop effects for low than high item-specific PC items (Bugg et al., 2011; Bugg & Hutchison, 2013; Jacoby et al., 2003). The different PC items are randomly intermixed and presented within the same block, and the global probability of congruent and incongruent trials is usually kept equal in that block (list-wide PC of 50%), so that this manipulation cannot imply a control modulation at the list level. Indeed, participants learn the conflict likelihood of the items, but they can use this item-specific information to impose attentional biases only after stimulus onset. The item-specific PC effect is thus a reactive modulation of control that, by operating on an item-by-item basis, is fast and flexible and is maximal when the item signals a high level of interference (Bugg, 2012, 2017; Bugg et al., 2011; Bugg & Hutchison, 2013).

Overall, PC manipulations are fundamental as they allow scholars to investigate what is referred to as adaptive control, that is, the context-induced and time-varying adjustments intrinsic to cognitive control (Braem et al., 2019). However, several authors have called into question the validity/purity of adaptive control measures, including the PC manipulations discussed so far, claiming that they suffer from methodological issues (see Braem et al., 2019; Schmidt, 2019 for reviews).

First, there is considerable controversy about whether cognitive control per se is necessary to resolve the Stroop interference elicited by PC manipulations or, conversely, whether adaptive control measures are merely the result of much simpler stimulus–stimulus or stimulus–response associative learning processes, as claimed by the contingency hypothesis (Schmidt, 2019; Schmidt & Besner, 2008; Schmidt et al., 2007). This view identifies contingency learning as a more plausible candidate for explaining PC effects, excluding the involvement of any high-level cognitive control modulation. Essentially, it postulates that by learning that responses tend to co-occur with specific stimuli, they can be facilitated. According to this hypothesis, participants' cognitive system implicitly learns the contingencies (or correlations) between the task-irrelevant and the task-relevant stimulus features, namely the responses, and uses the task-irrelevant dimension to predict high-contingency responses. When responses are highly predictable, namely, in high-contingency trials, participants exploit (even implicitly) these learning-based shortcuts and respond faster (Schmidt & Besner, 2008; Schmidt et al., 2007).

The contingency hypothesis challenges especially the item-specific PC effect, pointing out that such effect is only incidental, since item-specific PC manipulations are always confounded with contingency (Schmidt, 2019; Schmidt & Besner, 2008). Indeed, to manipulate the PC of the items, the frequency of specific irrelevant-relevant characteristic pairs is necessarily altered as well (Spinelli et al., 2019). In line with this hypothesis, the assessment of contingency learning controlling for PC effect (high- vs. low-contingency items of equal PC) revealed a contingency effect, while the assessment of PC effect controlling for contingency (high vs. low item-specific PC items of equal contingency) yielded no residual PC effect (Schmidt, 2013; Schmidt & Besner, 2008; see also Schmidt, 2019, for a detailed review). However, other evidence argues in favor of a more intermediate account that embraces the contribution of both contingency and item-specific control mechanisms. For example, in Bugg et al. (2011), the task-relevant dimension signaled the item-specific PC rather than the task-irrelevant one and this, by equating contingency across conditions, allowed deconfounding item-specific PC and contingency and finding evidence for a control modulation. When, instead, the task-irrelevant dimension functioned as the item-specific PC signal, contingency was confounded with item-specific PC and accounted for its effect, as predicted by the contingency hypothesis (Schmidt, 2019). Moreover, Bugg and Hutchison (2013, Experiment 3), restoring the traditional

[1] A third type of PC manipulation exists, the context-specific PC. Besides being akin to the item-specific PC, we will not discuss in detail the context-specific PC as it was not manipulated in the present work, but see Bugg and Crump (2012) and Bugg (2012) for reviews.

item-specific PC design (the task-irrelevant dimension signaling the item-specific PC), found an item-specific PC-dependent control modulation when four-item sets were used, while contingency was dominant when two-item sets were used, suggesting that bigger set sizes promote reliance on item-specific control, whereas smaller ones favor the use of contingencies. Their findings support the existence of different mechanisms governing the item-specific PC effect depending upon the set size, with larger sets reducing high-contingency responses and the likelihood of learning contingency associations, especially for responding to incongruent items.

Overall, this issue is still a matter of debate and a detailed discussion goes beyond the scope of the present work (see also Bugg, 2014; Bugg & Hutchison, 2013; Schmidt, 2013, 2019). However, what is clear is the need of methodologically correct/appropriate experimental designs controlling for or removing contingency-related biases to verify whether the item-specific PC effect is, even only partially, due to congruency modulation. To this end, apart from the strategy reported above, another solution is to manipulate contingency learning and item-specific PC in a partially independent way. When using a color-word Stroop task, this can be done by creating two nonoverlapping two-item-sized sets (but see Spinelli & Lupker, 2020 for a more detailed explanation, noting that, for congruent items, this does not dissociate between the two accounts). Recently, Braem et al. (2019) summarized some guidelines on how to design cognitive control tasks avoiding confounds and, for what concerns item-specific PC manipulations, they suggest creating two sets of overlapping "inducer" items (one for low and one for high item-specific PC items) to trigger reactive control and a third set of "diagnostic" items with a PC of 50%, to which the item-specific PC-dependent level of cognitive control is assumed to be transferred, to measure the item-specific PC effect without item-frequency differences. In Braem et al.'s (2019) consensus paper, different solutions are described in detail but it is worth mentioning that the authors admit that each of these solutions comes with cautionary notes. Indeed, they recommend one of the approaches but still highlight that it has been rarely tested, and that it is "important to assess its robustness in future studies" (p. 778).

Another potential flaw of adaptive control measures concerns the list-wide PC manipulation. Some authors indeed pointed out that list-wide PC effects do not actually depend on list-level information but instead can be explained by a mechanism operating at the item level. Thus, this account excludes any proactive control involvement or, at least, posits that it cannot be elicited by the list-wide PC manipulation (Blais & Bunge, 2010; Blais et al., 2007; Bugg et al., 2008). Indeed, when low list-wide PC blocks are composed of low item-specific PC items and high list-wide PC blocks are composed of high item-specific PC items, list-wide PC is confounded with item-specific PC. To disentangle the two mechanisms, Bugg and Chanani (2011) randomly intermixed, in both high and low list-wide PC blocks, an additional set of items with an item-specific PC of 50% (unbiased or diagnostic items) to verify whether a list-wide PC effect could be observed for such items, which did not provide any item-specific or contingency information. This was the case, suggesting that the list-wide PC effect was driven by a mechanism using the information at the list-level, and thus it was modulated proactively. This procedure was proposed by Braem et al. (2019) too, who agreed on the need to use inducer items that trigger proactive control and diagnostic (or unbiased) items that measure its effect on performance. They also recommended presenting inducer items more frequently than diagnostic ones and using a set of at least three items.

Although there is an emerging consensus on the need to use the approaches described above to design confound-minimized studies (Braem et al., 2019), their implementation comes at a cost. Indeed, both for list-wide PC and item-specific PC measures, distinguishing between inducer and diagnostic items is impractical and time-consuming due to the need to measure PC-related effects only on diagnostic items, while excluding inducer ones from the analyses. Moreover, for what concerns item-specific PC manipulation specifically, the creation of multiple sets of stimuli consisting of a multitude of items is not always feasible because, except for picture-word Stroop tasks, the possible exemplars of items are limited (e.g., for the color-word Stroop task there are just limited colors among which to choose). Lastly, there is no compelling evidence supporting the effectiveness of the transfer of item-specific PC from diagnostic to inducer items. Therefore, in our view, to date there is no methodological approach free from limitations and how to control for PC-related confounders in a feasible and effective way still remains an open question. Here, the solutions proposed by the confound-minimized approaches will be referred to as design-level control, as their purpose is to control for confounders as much as possible at the level of the experimental design. However, as just described, they imply some costs. An alternative to this approach is to control for confounding effects at the statistical level, for which we will use the label analysis-level control. This statistical approach, which we have adopted here (as described below), offers greater flexibility in the experimental design, thus overcoming the limitations of the approaches controlling for confounders at the design level.

These methodological controversies notwithstanding, the existence of two temporally distinct control modes seems plausible, at least as long as potential confounders are controlled for at the design level. However, the only way of verifying whether proactive and reactive control constitute truly independent mechanisms (Braver et al., 2007), ruling out that they are two poles on a continuum, is by obtaining independent estimates of these effects from the same sample of participants. This was done by Gonthier et al. (2016), who tried to dissociate proactive and reactive control by directly contrasting their behavioral signatures in a within-subject design to obtain independent estimates of list-wide PC and item-specific PC effects in the same participants. To this end, separate blocks were used: two list-wide PC blocks (one with low and one with high list-wide PC) along with a set of unbiased items to avoid item-specific, PC-related influence, and one item-specific PC block including an equal number of low and high item-specific PC items, with a list-wide PC of 50% to exclude list-wide PC effects. They found not only that list-wide PC and item-specific PC manipulations independently reduced the magnitude of the Stroop effect, but also that the two benefit indices were negatively correlated, suggesting that subjects relying more on one mechanism engage less the other one, thus providing evidence that the two effects are elicited by two distinct (i.e., dissociable) control mechanisms.

Although the study by Gonthier et al. (2016) provided initial evidence for the separability of list-wide PC and item-specific PC effects, thanks to its within-subjects design, and confirmed its suitability for measuring proactive and reactive control, it also suffers from some drawbacks. First, the effect of contingency learning was more strongly controlled for in the list-wide PC manipulation, for which unbiased diagnostic items were used while the influence of contingency on item-specific mechanisms was controlled for by using the task-relevant dimension to signal item-specific PC, leading

to unequal frequencies of unique trial types and irrelevant stimulus characteristics. A second limitation of the study by Gonthier et al. (2016) is that, although they use a within-subjects design testing both proactive and reactive control in the same participants, the two control mechanisms were investigated separately, as the list-wide PC and item-specific PC manipulations are kept apart and implemented one at a time in different blocks. Indeed, more convincing evidence for their existence as distinct mechanisms would require testing their interaction while both manipulations are implemented. Indeed, this would allow exploring whether and how they covary, informing about the existence of two separate mechanisms.

Aim of the Present Study and Methodological Novelties

Motivated by the considerations discussed above, our main aim here is to make a step further, by investigating in a more direct manner whether proactive and reactive control are two separable mechanisms. Essentially, we put forward a new approach that allows manipulating list-wide PC and item-specific PC at the same time, while controlling for the effect of stimulus-response associations, such as contingency. Although, to the best of our knowledge, only one attempt has been made to study both together using the Stroop task (see Hutchison, 2011), we believe that, to verify the specificity of these two control mechanisms, the most plausible way is to measure participants' performance while both list-wide PC and item-specific PC are parametrically varied at the same time. Indeed, by doing so, we can verify whether proactive and reactive control modes have distinct effects on participants' performance. Moreover, if we assume that they are indeed distinct mechanisms, they should also interact as (implicitly) predicted by the DMC model (Braver et al., 2009, 2021; De Pisapia & Braver, 2006). Thus, by measuring both at the same time, we can also test the three-way interaction between the Stroop effect, list-wide PC and item-specific PC, which can tell us more about the impact of variable amounts of proactive and reactive control activated by different levels of list-wide PC and item-specific PC. Lastly, since previous literature has confirmed the, at least partial, role of contingency on conflict adaptation, we decided not to exclude it but we allowed it to vary orthogonally (as much as possible) with respect to list-wide PC and item-specific PC, with the aim to estimate its effect and control for it at the statistical level so to measure the list-wide PC and item-specific PC effects regardless of contingency. As a consequence, in our tasks, trials will have at the same time a different level of list-wide PC, item-specific PC, and contingency. To do so, we combined this methodological approach with the use of a multilevel, trial-level modeling analytical approach to assess the fine-grained effects of our predictors at the subject level, while partialing out the effect of contingency and of other lower-level confounding factors (Viviani et al., 2024). As we recently showed, indeed, trial-level confounders represent important sources of trial-by-trial noise that cannot be accounted for by standard general linear models, which require collapsing trial-level data to obtain participants-by-condition averages. They can instead be effectively estimated and removed by multilevel modeling (see Viviani et al., 2024 for a more exhaustive description of the advantages of multilevel modeling over general linear models).

Moreover, to pursue our aim, we introduced an important methodological novelty, that is, we manipulated the different PCs at the list-wide and item-specific levels to explore how and to what degree they modulate the Stroop effect dynamically and in a fine-grained way. In other words, we aimed to evaluate the impact of PC on participants' Stroop performance on a trial-by-trial basis, using trial-level, list-wide PC and item-specific PC estimates computed based on the actual recent history of trial congruency they experienced, rather than on their assumed (future) experience of trial congruency at the block level.

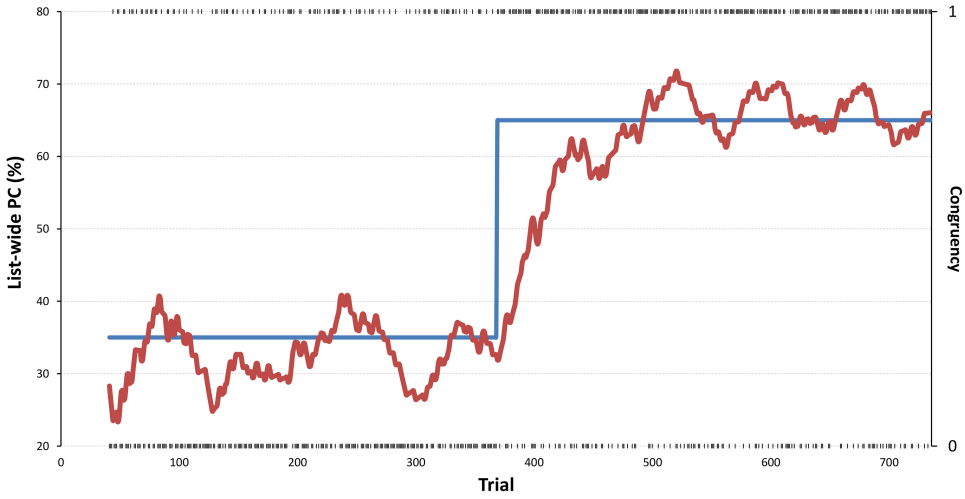
Indeed, it should be noted that the available literature used the block-level, list-wide PC and item-specific PC variables, that is, those computed as the number of congruent trials within a block (in total or for each item, respectively) divided by the total number of trials (in total or for each item, respectively) within the same block. However, these block-level, list-wide PC and item-specific PC variables correspond for sure to the actual PCs at the end of the block only. This is true even if trial-level PC values are computed based on the trials experienced so far during that block, and especially if they are computed based on the local history of trials (e.g., using a moving window or a forgetting factor). Indeed, due to the commonly used (pseudo)randomization of the trial list, it is not unlikely that the list-wide PC value at, say, the 20th trial in a block deviates even dramatically from the expected block-level, list-wide PC, being it, for example, as large as 40% and as small as 0% instead of 20%. This is especially important after an unsignaled block transition, especially between blocks with extreme opposite block-level, list-wide PCs (e.g., 20% and 80%). In this case, indeed, the commonly used block-level approach implausibly assumes that, at the very first trial of a new block, participants immediately update their list-wide PC estimates (from 20% to 80% in this example) and, consequently, their proactive control level. Similarly, commonly used block-level, item-specific PC values neglect the fact that participants first need to experience a sufficient number of trials for each item to estimate its item-specific PC value, thus unrealistically assuming that the items at the beginning of each block have already been associated to an item-specific PC value, without previously encountering them. Moreover, the commonly used block-level approach unrealistically assumes that all the trials within a block share the same PC values, not taking into account the fact that the local PCs vary within the block.

Therefore, and since participants are not aware of the probabilistic structure of the task, it is unreasonable to assume that their trial-by-trial performance is modulated by block-level PC values. Instead, it is more plausible to assume, as we do here, that their cognitive system implicitly and continuously estimates trial-level, list-wide PC and item-specific PC values using some form of statistical learning based on the recent history of overall and item-specific PCs, respectively, implementing a specific level of control accordingly. We therefore employed a fine-grained manipulation of list-wide PC and item-specific PC, which were estimated trial-by-trial using an ideal Bayesian observer (Mathys et al., 2011). Our approach, thus, allows us to account for and estimate flexible, ongoing adjustments of cognitive control during the task (see Figure 1). Trial-by-trial estimates (which we will call continuous variables) were used as predictors in our analyses as they are more realistic than those computed using the block-level occurrences (which we will call discrete variables). Trial-level estimates were also calculated for confounding variables of interest, including contingency, using the same approach.

Finally, we addressed the important but frequently overlooked aspect of measure reliability. As highlighted by Gonthier et al. (2016), list-wide PC and item-specific PC effect indices have unknown psychometric properties, in addition to being effects

Figure 1

List-Wide PC Estimates



Note. The plot shows the block-level, list-wide pc (blue [light gray] line) and its trial-level estimates (red [dark gray] line) computed using the hierarchical Gaussian filter (Mathys et al., 2011) for one of the trial lists used in the experiment. The occurrence of congruent (congruency = 1) and incongruent (congruency = 0) trials is also depicted as small black vertical markers. PC = proportion congruency. See the online article for the color version of this figure.

calculated from difference scores, which, in turn, further reduces their reliability (Thomas & Zumbo, 2012). Despite this awareness, this issue has rarely been addressed in studies using such manipulations, and, as such, our study also aims to explore the reliability of such measures. To this aim, the use of multilevel, trial-level modeling of participants' performance is again fundamental, as we recently showed it to ensure estimations of the experimental effects with higher and more stable internal reliability compared to standard general linear model approaches (see Viviani et al., 2024).

The three points mentioned above were addressed in two experiments involving four-choice spatial Stroop tasks that require keypress responses to indicate the direction of a target arrow, ignoring its position. Both experiments used exactly the same experimental procedure and design and differed only in the spatial arrangement of the experimental stimuli (see below), allowing us to assess the robustness of our experimental approach and results. These two spatial Stroop tasks, named peripheral and perifoveal spatial Stroop tasks, were chosen as they overcome some limitations intrinsic to the original color-word verbal Stroop task, while also ensuring a complete Stroop effect, that is, an effect including conflict at the task, stimulus, and response levels (see Viviani et al., 2023 for more details). In addition to these methodological advantages, in a recent work, we have shown that the peripheral and perifoveal spatial Stroop tasks are proper spatial Stroop adaptations, producing Stroop effects that not only have a large magnitude but are also robust to analytical flexibility and have a high and robust internal reliability (Viviani et al., 2024).

Hypotheses

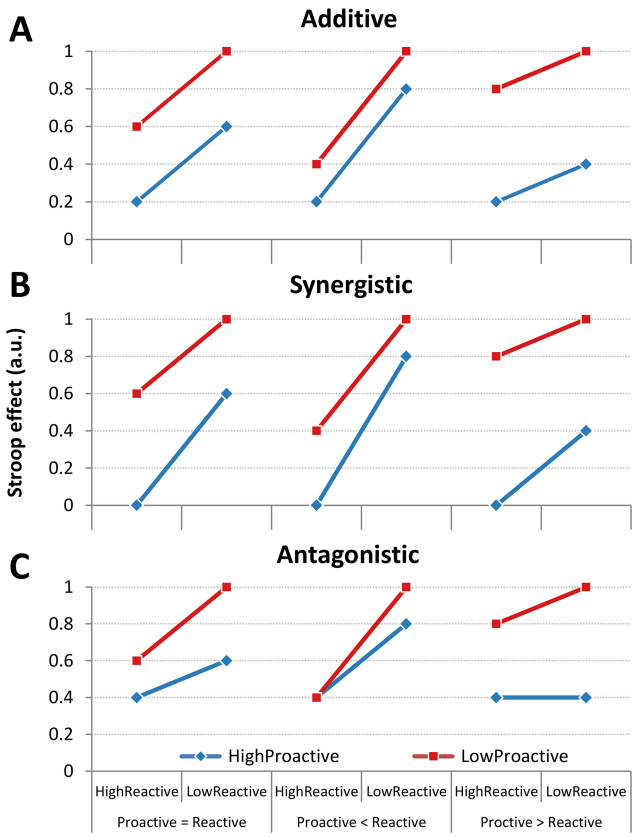
As claimed above, we were interested in exploring if and how proactive and reactive control covary and interact to modulate the Stroop effect. To the best of our knowledge, this interaction has rarely been tested before, mainly because list-wide PC and item-specific PC have always been manipulated separately (but see Hutchison, 2011). As such, there is no solid evidence of how the Stroop effect is modulated when both proactive and reactive controls are implemented in the same experimental design and neither of whether these two control modes interact. Therefore, we put forward some hypotheses about what we expect to observe, proposing different theoretically plausible alternatives.

All the hypotheses assume that when both list-wide PC and item-specific PC are high (LowProactive/LowReactive condition), the lowest level of control is applied (i.e., no form of proactive and reactive control is implemented) and thus the Stroop effect should be the largest (equal to 1 in our models). Conversely, when either list-wide PC or item-specific PC are low, a high level of proactive and reactive control, respectively, should be implemented (respectively, the HighProactive/LowReactive and HighReactive/LowProactive conditions), and thus the Stroop effect should be reduced. Finally, when both list-wide PC and item-specific PC are low (HighProactive/HighReactive condition), the highest level of control should be implemented and thus the smallest Stroop effect should be observed.

The first point that differentiates our hypotheses is the size of the Stroop effect in the HighProactive/HighReactive condition. We hypothesized that, if proactive and reactive control do not interact with each other but still separately modulate the Stroop effect, their effects on the Stroop effect will be additive, thus still producing the smallest Stroop effect compared to the other conditions (additive models; Figure 2A). If, in contrast, their interaction is significant, two alternative scenarios are possible: They could interact either in a synergistic (i.e., more than additive) way, producing a reduction of the Stroop effect that is greater than that assumed by the additive hypothesis (synergistic models; Figure 2B), or in an antagonistic (i.e., less than additive) way, producing a reduction of the Stroop effect that is smaller than that assumed by the additive hypothesis (antagonistic models; Figure 2C).

Figure 2

Predicted Patterns of Stroop Effect Modulations



Note. Predicted patterns of Stroop effect modulations by low and high levels of list-wide PC-related proactive control (proactive) and item-specific, PC-related reactive control (reactive) according to our alternative hypotheses. Additive models, additive effects of proactive and reactive; synergistic models, synergistic proactive by reactive interaction; antagonistic models, antagonistic proactive by reactive interaction (see Hypotheses section for more details). To note that HighProactive and HighReactive conditions correspond to low list-wide and item-specific PCs, respectively; LowProactive and LowReactive conditions correspond to high list-wide and item-specific PCs, respectively. PC = proportion congruency; a.u. = arbitrary unit. See the online article for the color version of this figure.

The second distinction stems from the possibility that one of the two control modes could have a stronger impact on the Stroop effect than the other. This point differentiates our hypotheses only for what concerns the conditions wherein only proactive control or reactive control is implemented (respectively, HighProactive/LowReactive and LowProactive/HighReactive), while it should not affect the size of the Stroop effect in the conditions wherein neither or both forms of control are implemented (LowProactive/LowReactive and HighProactive/HighReactive, respectively). If we assume that proactive and reactive controls have the same strength, the Stroop effect should be the same size in the LowProactive/LowReactive and HighProactive/HighReactive conditions (Figure 2, left plots). Conversely, if we assume that the effect of proactive control is stronger, the Stroop effect should be smaller in the HighProactive/LowReactive condition compared to the LowProactive/HighReactive condition (Figure 2, right plots), while if the effect of reactive control is stronger, the Stroop effect should be smaller in the LowProactive/HighReactive condition compared to the HighProactive/LowReactive condition (Figure 2, middle plots).

Experiment 1—Peripheral

Method

Procedure and Experimental Tasks

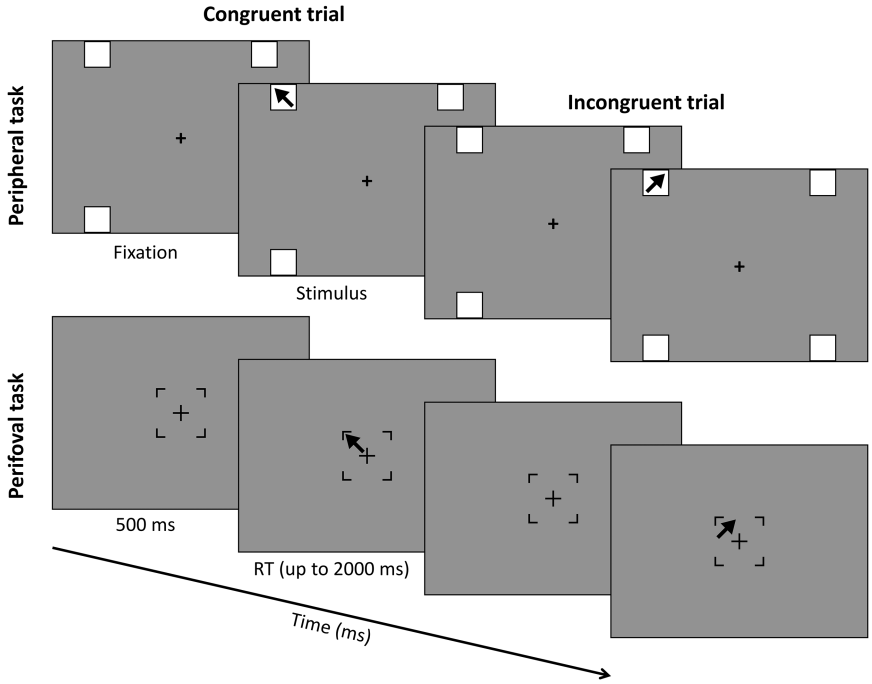
The experiment was programmed using Psytoolkit (Stoet, 2010, 2017) and administered online. The stimuli were presented in full-screen mode, with a resolution of 800 × 600 pixels, on a gray background (red, green, and blue: 128, 128, 128). Each trial started with a fixation stimulus presented at the center of the screen for 500 ms and participants were instructed to fixate it. Then, the experimental stimulus appeared and remained on the screen until participants responded or up to a response time-out of 2,000 ms (Figure 3). Participants had to pay attention to the task-relevant information, which was the pointing direction of a black arrow, and were required to indicate it via button press by using four keys on a computer keyboard, which were E, O, K, and D. These keys were spatially arranged to be compatible with the four possible arrow directions, which could be upper left, upper right, lower right, or lower left, and had to be pressed using the left middle, right middle, right index and left index fingers, respectively. The experimental stimuli were also characterized by task-irrelevant information, which was the position where the arrow appeared. The position of the arrow overlapped with the four task-relevant directions, since the arrow could appear in an upper-left, upper-right, lower-right, or lower-left position. The task-irrelevant position could match or not the task-relevant direction, yielding congruent and incongruent trials, respectively.

In this study, we used a peripheral spatial Stroop task (Viviani et al., 2024), wherein the target arrow could appear in one of four peripherally located spatial positions. For this task, the fixation screen consisted of a black cross (36 × 36 pixels) presented at the center of the screen, along with four white squares (73 × 73 pixels) at the four corners of an imaginary square of 600 × 600 pixels centered on the screen. Then, the target arrow was presented inside one of the four peripheral squares, and it could point to one of the same four directions (Figure 3). We used 12 of the 16 possible combinations of arrow positions and directions, as we excluded the four corresponding to the incongruent arrows pointing to the opposite direction (e.g., the arrow appearing at the upper-left corner and pointing toward the bottom-right corner) because they point toward the correct response.

List-wide PC and item-specific PC proportions of congruency were simultaneously manipulated to measure both proactive and reactive control, respectively. To this aim, the trial lists were designed by first dividing them into two main blocks made of 320 experimental trials each, with different list-wide PC values, one with 35% of congruent trials (list-wide PC 35) and one with 65% of congruent trials (list-wide PC 65). Then, distinct item-specific PC levels were used, nested within each list-wide PC block, to have four different item-specific PC values within each block. In the list-wide PC 35 block, the item-specific PC ranged from 20% to 50% in steps of 10%, while in the list-wide PC 65 block, the item-specific PC ranged from 50% to 80% in steps of 10%. The item-specific PC values of the items were deliberately flipped between

Figure 3

Temporal Structure of the Peripheral and Perifoveal Tasks



Note. For each task, exemplar congruent and incongruent trials are depicted. RT = response times.

the two list-wide PC blocks to enhance the trial-level, item-specific PC variability between trials and to improve the estimation of its effect (see below). Crucially, by using the same item-specific PC level (50%) in both list-wide PC blocks, we were able to assess the pure effect of list-wide PC (and thus proactive control) on Stroop effects, independently of item-specific PC (and thus reactive control) and contingency. Moreover, within each block, the occurrence of each position-direction combination was intentionally varied in trying to orthogonalize as much as possible the contingency to list-wide PC and item-specific PC, so that the effect of each of these variables could be disambiguated in the statistical analysis. In doing so, we allowed the probability of each of the four directions (and thus the responses) to slightly vary within each subblock, while keeping the probability of each of the four positions constant. We thus obtained different contingency values, ranging from 5% to 80%, and within each item-specific PC level, two different contingency values differing by 10% were used for the two possible incongruent trials (see Figure 4). The obtained correlations between the contingency and the list-wide PC and item-specific PC variables were, respectively, $r = .393$ and .483, corresponding to 15.4% and 23.4% of shared variance. Of note, the nonzero correlation between contingency and item-specific PC is due to the fact that, for congruent trials, they are perfectly correlated because they are computed in the same way and their occurrences are the same (see Figure 4).

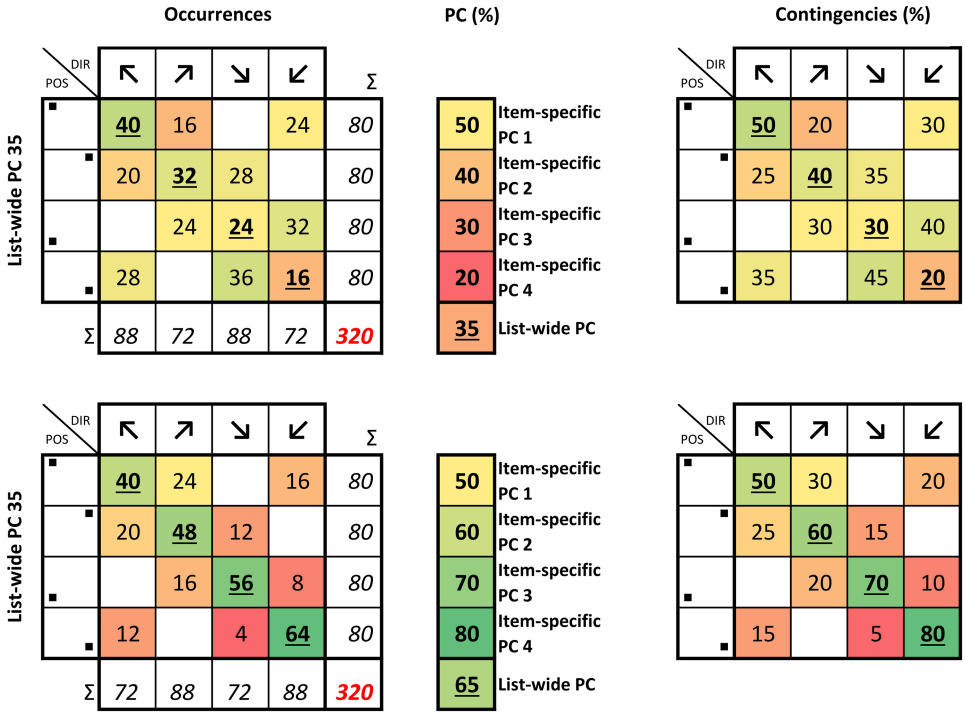
We created 16 lists which were randomly assigned to participants. In these 16 lists item-specific PC and contingency were counterbalanced across items. The fact that item-specific PC values were counterbalanced across lists is crucial to ensure that the effect of item-specific PC did not depend on specific items but truly operated in the expected direction, given that item-specific PC values were flipped between the two list-wide PC blocks. Additionally, the statistical model we utilized (see below) accounted for item-specific PC values regardless of the specific item. To further ensure that the manipulation of item-specific PC worked as expected, considering results repeatedly showing its nonsignificant effect, we conducted post hoc control analyses, testing the effect of item-specific PC separately in the two list-wide PC blocks. These analyses were conducted on block-level, item-specific PC values, comparing the effect of item-specific PC values on the Stroop effect in the two blocks (item-specific PC 20 vs. item-specific PC 50 in the list-wide PC 35 block and item-specific PC 80 and item-specific PC 50 in the list-wide PC 65 block). They confirmed that the item-specific PC manipulation worked as expected, as the Stroop effect increased with higher item-specific PC values in both blocks, $ts(97) = 5.05$ and 3.30, $ps < .002$, $ds = 0.51$ and 0.33, respectively.

In addition to the 640 experimental trials, before each list-wide PC block, we added subblocks of 40 trials to favor the familiarization of participants to the current block list-wide PC level. Moreover, the 640 experimental trials were divided into eight subblocks with self-paced breaks in between, and at the beginning of each subblock we added two buffer trials. The habituation and buffer trials were then excluded from the analyses. Within each trial-list, the order of presentation of the trials was pseudorandomized using the software Mix (van Casteren & Davis, 2006) to avoid more than four consecutive repetitions of the same congruency and both total and partial repetitions of stimulus characteristics and/or responses in order to control for first-order priming effects.

A second step in trial-list design was to compute trial-wise, list-wide PC, item-specific PC, and contingency for each trial-list version using the hierarchical Gaussian filter (Mathys et al., 2011). The

Figure 4

Block-Wise Structure of the Task



Note. Separately for one subblock of each list-wide PC block (list-wide PC 35 block, top row; list-wide PC 65 block, bottom row), the image shows on the left the number of trials (occurrences) with a specific target direction and position. For example, in the list-wide PC 35 subblock, we had nine trials with the arrow appearing in the lower-left corner, but pointing toward the lower-right corner. The trials in the diagonal are the congruent ones (underscored). For each subblock, the corresponding contingencies are also shown on the right, while the middle column shows the percentage of congruent trials (list-wide PC) and of congruent trials specific for each location (item-specific PC). The color scale indicates the relative probability of each trial type/contingency, as well as the relative level of the list-wide PCs/item-specific PCs. DIR = direction; POS = position; PC = proportion congruency. See the online article for the color version of this figure.

hierarchical Gaussian filter enables estimating trial-wise trajectories of beliefs about the hidden environmental states (e.g., the actual list-wide congruency) that caused task inputs (e.g., trial-wise congruency). The hierarchical Gaussian filter presents a series of advantages that converted it into a largely used toolbox for the empirical investigation of learning under uncertainty, also in the context of spatial Stroop tasks (Visalli et al., 2023). First, hierarchical Gaussian filter adopts a variational approximation to Bayesian inference that—avoiding the complexity of exact Bayesian inference computations—can account for real time learning in a biologically plausible fashion. Moreover, in the hierarchical Gaussian filter the inferred states are hierarchically organized and evolve in time as Gaussian random walks with step sizes (volatility or variance) determined by the next highest hierarchical level. In this way the learning rate is not assumed to be fixed during the task, but it flexibly adapts to the environmental volatility estimated at higher levels (Behrens et al., 2007). Specifically, trial-level estimates were computed reflecting trial-wise probabilities updated based on: (a) the stimulus congruency, for list-wide PC; (b) the stimulus congruency conditional to a specific position, for item-specific PC; (c) the target direction (and thus the response) conditional to a specific position, for contingency. As an additional advantage of estimating the trial-wise variables, compared to using the block-level ones, the trial-wise, list-wide PC and item-specific PC were better orthogonalized to the trial-wise contingency. Indeed, their collinearity (which is inevitably high in the block-level variables, as explained above) was decreased to, respectively, $r = .304$ and .360 (corresponding to 9.3% and 12.9% of shared variance). The hierarchical Gaussian filter was also used to compute trial-by-trial probabilities of other variables used as confounding predictors in statistical analyses (see below).

Before beginning the task, the participants received general instructions on the procedure, the task, and the response mapping. Considering also that the task was completed online, we took particular care to keep the instructions as simple, detailed, and clear as possible. Participants were asked to respond as quickly and accurately as possible and recommended performing the task in a quiet environment, maintaining a comfortable posture, and keeping the responding fingers in contact with the response keys. After the instructions, the participants completed a block of 20 practice trials with list-wide PC and item-specific PC at 50%, during which they received feedback on their performance, and, in the case of errors or time-out responses, they were also provided with a brief summary of instructions and response mapping. Practice trials were presented until 75% accuracy was reached.

Data Analysis

Various analyses were conducted to estimate the effect of list-wide PC and item-specific PC manipulations on the magnitude of the Stroop effect, while controlling for contingency, and to estimate the internal reliability of our effects. Statistical analyses were conducted using Matlab (Version 2017b; The MathWorks, Inc. Natick, Massachusetts, United States).

The analyses were performed on inverse-transformed RTs (iRTs, computed as −1,000/RTs). This transformation was employed to eliminate the heavy right skewness of the RT distribution, for which the logarithmic transformation was not enough. From the 62,720 experimental trials, we excluded trials with incorrect responses ($n = 2{,}115$, corresponding to 3.37% of the experimental trials), missed responses ($n = 235$, corresponding to 0.37% of the experimental trials), and RTs shorter than 150 ms ($n = 1$, corresponding to <0.01% of the experimental trials), which were all treated as errors, and post-error trials ($n = 2{,}103$, corresponding to 3.35% of the experimental trials). Control analyses were performed on both untransformed RTs and natural log-transformed RTs (lnRTs) to assess the robustness of the results to analytical flexibility.

We checked for the presence of participants with low compliance, defined as those having either a mean iRTs more than three standard deviations away from the sample mean or a mean accuracy lower than 70% (i.e., the level used in the practice block). Based on these criteria, no participant was excluded from the analyses (see Figure S1.2 in the additional online materials at https://osf.io/aeh2d).

Assessing the Magnitude of List-Wide and Item-Specific PC Effects. The statistical analyses were performed using a multilevel modeling approach, also called trial-by-trial hierarchical modeling, by performing linear mixed-effects model analyses (Baayen et al., 2008). This approach is the most suitable for our experimental design and, thus, our aims. Indeed, using multilevel modeling, we were able to assess the distinct impact of list-wide PC and item-specific PC, as well as their interaction, in modulating the Stroop effect when they both varied. Moreover, this approach allowed us to do so while partialing out the effect of contingency and other lower-level confounding factors, which represent sources of trial-by-trial noise in the estimation of our effects of interest at the subject level. Finally, this approach allowed us to employ the trial-level estimates of our predictors because, as explained in the introductory part, considering trial-by-trial history is more realistic than using the respective discrete values. Multilevel modeling also allows one to overcome standard general linear model drawbacks. Indeed, we recently showed that this approach ensures more accurate and precise estimates of the experimental effects of interest. Moreover, since this approach explains intrasubject/intertrial sources of variance contributing to measurement error, it also provides better reliability of these estimates (Viviani et al., 2024).

We tested a linear mixed model defined a priori based on our theoretical assumptions, which we named "continuous full model." Indeed, this model included the trial-level estimates of our predictors for both experimental manipulations and confounders (we will indicate those predictors with "trial level" before their names). Specifically, in the fixed part of the model, we included several confounding predictors, for the reasons explained above. Each confounder was included based on well-known effects in the literature. Specifically, we included (a) a continuous predictor reflecting the iRT of the preceding trial (iRTpreceding), to account for temporal dependency in RTs (Baayen & Milin, 2010) and thus to avoid violating the assumption of the independence of observations for linear modeling; (b) the continuous predictor for the effect of trial number (trial) to account for potential time-on-task effects, such as the effects of learning/adaptation or fatigue; (c) the horizontal and vertical position of the stimulus on the screen (respectively, HorizontalStimulus and VerticalStimulus), to account for potential (e.g., perceptual, attentional) differences due to the location where the stimulus appeared (left vs. right, above vs. below, respectively); (d) the horizontal and vertical coding of the response (respectively, HorizontalResponse and VerticalResponse), to account for potential (e.g., motor) differences due to the response hand and finger, respectively. Lastly, we included predictors for low-level learning effects that have been shown to affect the Stroop interference resolution, threatening the interpretability of the Stroop performance with control-related accounts, that is, (e) contingency, which is the conditional probability of the response given the stimulus, P(R|S), and (f) the probability of the response, P(R) (ResponseProbability). The experimental effects of interest were modeled by including the predictors for the list-wide PC and item-specific PC manipulations (list-wide PC and item-specific PC, respectively) and trial congruency (congruency), as well as their interactions. The three-way interaction served to explore whether proactive and reactive control interacted in modulating the Stroop effect, and it was included in both the fixed and random parts of the model, as we assumed that it varied across subjects. The Wilkinson-notation formula for the continuous full model is as follows: RT ~ iRTpreceding + Trial + HorizontalStimulus + VerticalStimulus + HorizontalResponse + VerticalResponse + Trial-Level Contingency + Trial-Level Response Probability + Trial-Level, List-Wide PC × Trial-Level, Item-Specific PC × Congruency + (Trial-Level, List-Wide PC × Trial-Level, Item-Specific PC × Congruency|Subject).

The continuous predictors iRTpreceding, trial, trial-level contingency, and trial-level ResponseProbability were centered and scaled at the participant level to facilitate the convergence of the model and the interpretation of the results, while scaling was not necessary for trial-level, list-wide PC and trial-level, item-specific PC, since, by calculating their trial-level estimates, they were already on a scale centered at a 50% probability. The predictor for congruency was coded with the values of 0 and 1 for the incongruent and congruent conditions, respectively, with the latter acting as the reference level.

Then, to assess whether there was evidence of stress in the model fit, after fitting the model, we inspected the model residuals and we then refitted a trimmed version of the model obtained by excluding data points with absolute standardized residuals exceeding 3.

We report the estimated coefficient (b), SE, and t and p values for each fixed effect included in the trimmed final model. We calculated the p-values by using Satterthwaite's approximation of degrees of freedom, which was also used to compute the corresponding effect size estimates (d_S) for the experimental effects of interest. The effect sizes for the same effects were also computed as standardized differences based on the participants' estimated condition means based on their random slopes (d_r). An α level of .05 was set as the cutoff for statistical significance. We used the participants' random slopes to compute the individual effects of interest (i.e., the Stroop effect and its modulation by list-wide PC, item-specific PC, and their interaction) and the corresponding dominance values, that is, the percentage of participants showing them.

We also performed some control analyses to verify whether our continuous full model was justified and ensured the best fit to the data. First, the same model was tested also using the block-level estimates of our variables, referred to as discrete variables (discrete full model), to assess the assumed theoretical advantages of the trial-level estimates (besides favoring comparability with previous literature). To this aim, we compared the Akaike information criterion (AIC) of the continuous and discrete full models to assess which one better explained our data. Moreover, to verify whether the inclusion of confounders actually increased the model goodness of fit, we compared the continuous full model to a reduced one (continuous reduced model), which included only the variables of theoretical interest but none of the confounding ones, using the log-likelihood ratio test (Baayen et al., 2008). Lastly, in the case in which the three-way interaction was not significant, we tested the same continuous full model but after removing the term for the three-way interaction (i.e., leaving the terms for the two two-way interactions congruency: trial-level, list-wide PC and congruency:trial-level, item-specific PC), to verify whether its inclusion might have interfered with the estimation of the effects of the two two-way interactions testing for the distinct effects of proactive and reactive control (continuous full two-way model). See Table 1 for the predictors included in the various models we tested, including the control ones.

To assess the robustness of our results to analytical flexibility, control analyses were also performed by replicating linear mixed model results for the continuous full model using another multilevel modeling approach, that is, a random coefficient analysis (also called random regression or two-step regression; Lorch & Myers, 1990). For the random coefficient analysis, we first ran linear regressions at the subject level using the same regression model as the final linear mixed model described above (continuous full model). The Wilkinson-notation formula for the random coefficient analysis model is: RT ~ iRTpreceding + Trial + HorizontalStimulus + VerticalStimulus + HorizontalResponse + VerticalResponse + Trial-Level Contingency + Trial-Level ResponseProbability + Trial-Level, List-Wide PC × Trial-Level, Item-Specific PC × Congruency.

As for the linear mixed model analysis, the model was refitted after the exclusion of data points with standardized residuals exceeding 3. Then, we assess the statistical significance and effect size of the tested effects at the group level by performing two-tailed, one-sample t tests against 0 on the estimated b coefficients for each participant.

We also performed additional analyses to assess the magnitude of the Stroop effect using a general linear model approach, which is the standard approach in cognitive psychology and relies on the aggregation of the participants' performance in trials of different conditions to obtain participants-by-condition scores. However, this approach discards any trial-by-trial variability that can contaminate participant-by-condition scores, potentially decreasing their accuracy and generalizability (Rouder & Haaf, 2019). More importantly, general linear model analyses are not well-suited for our experimental design because it is incomplete (item-specific PC is nested in list-wide PC). This was not an issue for testing the Stroop effect magnitude per se, as we could aggregate congruent and incongruent trials across the list-wide PC and item-specific PC levels, but it prevented us from testing the effects of list-wide PC and item-specific PC while also controlling for contingency (see assessing the magnitude of list-wide PC and item-specific PC effects). Indeed, due to our manipulation of list-wide PC, item-specific PC, and contingency, and since contingency is inevitably confounded with item-specific PC in congruent trials (because they are computed in the same way), we did not have all the required combinations of list-wide PC, item-specific PC, and contingency levels, and the trial number for the available combinations was heavily unbalanced. These limitations notwithstanding, we decided to perform general linear model analyses anyway to favor the comparison of our results on the Stroop effect magnitude with the literature.

Assessing the Internal Reliability of List-Wide PC and Item-Specific PC Effects. The internal consistency of the

Table 1
Tested Models and the Included Predictors

Predictors	Continuous full	Continuous full two-way	Continuous full NoContingency	Continuous reduced	Discrete full	Continuous full control
Intercept	X	X	X	X	X	X
Trial	X	X	X		X	X
Congruency	X	X	X	X	X	X
iRTpreceding	X	X			X	X
HorizontalStimulus	X	X	X		X	X
VerticalStimulus	X	X	X		X	X
HorizontalResponse	X	X	X		X	X
VerticalResponse	X	X	X		X	X
List-wide PC	X[a]	X[a]	X[a]	X[a]	X[b]	X[a]
Item-specific PC	X[a]	X[a]	X[a]	X[a]	X[b]	X[a]
Contingency	X[a]	X[a]		X[a]	X[b]	X[a]
ResponseProbability	X[a]	X[a]	X[a]	X[a]	X[b]	X[a]
Congruency:list-wide PC	X[a]	X[a]	X[a]	X[a]	X[b]	X[a]
Congruency:item-specific PC	X[a]	X[a]	X[a]	X[a]	X[b]	X[a]
List-wide PC:item-specific PC	X[a]	X[a]	X[a]	X[a]	X[b]	X[a]
Congruency:list-wide PC:item-specific PC	X[a]		X[a]	X[a]	X[b]	X[a]
Congruency:list-wide PC:contingency						X[a]

Note. Predictors included in each model are indicated by X. PC = proportion congruency; iRT = inverse-transformed response times.
[a] These effects refer to trial-level variables as explained in Assessing the Magnitude of List-Wide and Item-Specific PC Effects section. [b] These effects refer to block-level variables as explained in Assessing the Magnitude of List-Wide and Item-Specific PC Effects section.

experimental effects of interest was assessed for the linear mixed model results by computing split-half Pearson's correlations corrected with the Spearman–Brown formula (r_{SB}). We used 2,000 randomizations and calculated both the median r_{SB} values and the corresponding nonparametric 95% confidence intervals (95% CIs).

Essentially, observations were randomly split into two subsets and, on each subset, linear mixed model analysis was performed. As highlighted above, this allowed us to model the interindividual variability in the effects of interest (Stroop effect, proactive and reactive control, and their interaction), while partialing out the effect of the confounding predictors described above. Then, the by-subject random slopes for the effects of interest in the two subsets were correlated to obtain the r_{SB} values.

Participants

For the first experiment, we recruited 98 participants (55 female and 43 male participants; $M_{age} = 25.89$ years, $SD = 6.42$ years). Participants reported demographic information by completing free-response boxes for age gender (the labels "male," "female," or "other/prefer to not respond" were provided). No other demographic information was collected. Participants' handedness was assessed using the Edinburgh Handedness Inventory (EHI; Oldfield, 1971). The sample comprised five left-handed participants (EHI scores < -50) and nine ambidextrous participants (EHI scores between -50 and 50). No participants reported suffering from neurological or psychiatric disorders or being under medication. Participants gave their informed consent to participate in the study, which was conducted in accordance with the ethical standards of the 2013 Declaration of Helsinki for human studies of the World Medical Association. The study was approved by the Ethical Committee for the Psychological Research of the University of Padova.

Participants consisted of a convenience sample recruited using researchers' personal networks and were not compensated for their participation. To determine the sample size for the linear mixed model analysis, the approaches available to date for power analysis are not adequate and/or feasible for our complex statistical model (see Viviani et al., 2024 for a detailed discussion), especially because it involves the interaction between continuous predictors. Nonetheless, it should be noted that the random coefficient analysis and linear mixed model approaches are quite similar and provide similar results (at least regarding the Stroop effects in our experimental paradigm), and the power analysis for random coefficient analysis is trivial, as it concerns a simple one-sample *t* test on the by-subject slopes for the effect of interest. We thus performed an a priori power analysis in G*Power 3 (Faul et al., 2007) to compute the minimum sample size required to detect, with a statistical power of .80, the effect of main interest (i.e., the three-way interaction reflecting the Stroop effect modulation by the interaction between list-wide PC and item-specific PC) in a two-tailed, one-sample *t* test. We conservatively assumed a small-medium Cohen's *d* effect size of 0.35. This analysis revealed that at least 67 participants were required. We nonetheless decided to recruit as many participants as possible, exceeding the required sample size, to be able to detect even smaller effects (by increasing the statistical power of our analyses) and to increase the precision of the experimental effects estimates. It is important here to note that linear mixed models tend to provide higher power than standard general linear model approaches like the one-sample *t* test we used here.

Transparency and Openness

We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, all manipulations, and all measures in the study. All inclusion/exclusion criteria were established prior to data analysis. Although the current work was not preregistered, it meets the transparency and openness promotion guidelines suggested by the journal. All data and materials, as well as the code used to run the experimental tasks and generate and analyze the data of the current study, are available in our project repository on the Open Science Framework (OSF) platform at https://osf.io/qmu7g. The available study materials, anonymized raw data, and codes are reusable by other researchers.

Results

Magnitude of List-Wide and Item-Specific PC Effects

For all the analyses, we report here only the results for iRTs. Indeed, as mentioned above, the distribution of RTs was heavily right-skewed and the residuals of the analyses on both RTs and lnRTs violated the assumptions of normality and homoscedasticity (see Figures S1.1 and S1.3–S1.5 in the additional online materials at https://osf.io/aeh2d).

General linear model-based analyses were performed using *t* tests. The overall Stroop effect (i.e., collapsing across list-wide PC and item-specific PC values) was significant, $t(97) = 35.50$, $p < .0001$, and with a very large effect size ($d = 3.59$). Our result indicates that all our participants were significantly slower in responding to incongruent as compared to congruent trials (dominance = 100%; see Table S1.1 in the additional online materials at https://osf.io/aeh2d).

Regarding the linear mixed model analysis, we first compared the continuous full model to the discrete full model using the AIC model selection and we found that the best-fit model was the continuous full model (AIC = 8,352 vs. 8,436.7 of the discrete full model). Then, we compared our continuous full model with the continuous reduced one (see Table 1) by performing the log-likelihood ratio test, which revealed that the continuous full model was justified, $\chi^2(7) = 7{,}530.5$, $p < .0001$, confirming that the inclusion of confounders increased the model fit. As such, here we report the results of the analysis performed on iRTs using the full model with continuous variables.

The conditional R^2 of the linear mixed model was 0.69% and 0.84% of the observations was removed as outliers (>3 absolute standardized residuals) to mitigate the stress of the model fit (i.e., to improve the normality of the residuals, see Figure S1.3 in the additional online materials at https://osf.io/aeh2d). This analysis revealed that all the confounding predictors were significant in modulating participants' iRTs (all $ps < .0001$; see Table 2). Specifically, our results suggest that participants were faster as the trials progressed and when they responded to stimuli appearing on the upper and right sides of the screen using the middle finger and the right hand. Moreover, we found a significant temporal dependency in iRTs (i.e., a positive correlation between iRTs at the current and preceding trial). Lastly, responses were faster when the probability of the response (trial-level ResponseProbability) was higher.

For what concerns our predictors of interest, we found that the Stroop effect (congruency) was significant ($p < .0001$), with slower responses to incongruent trials. The Stroop effect had a very large effect size ($d_r = 3.67$, $d_S = 2.13$) and a dominance value of 100%, that is, all participants showed a positive Stroop effect. The Stroop

Table 2
Results of the Linear Mixed Model Analysis for Experiment 1 (Continuous Full Model)

Effect	*b*	*SE*	*t*	*df*	*p*
Intercept	−1.9439	0.0350	−55.559	145.60	<.0001
Trial	−0.0923	0.0023	−40.030	1,967.43	<.0001
Congruency	0.3521	0.0123	28.620	179.86	<.0001
iRTpreceding	0.0551	0.0012	47.190	57,392.28	<.0001
HorizontalStimulus	−0.0104	0.0028	−3.700	12,850.75	.0002
VerticalStimulus	−0.0791	0.0027	−29.056	19,753.88	<.0001
HorizontalResponse	−0.0351	0.0026	−13.381	34,201.64	<.0001
VerticalResponse	−0.1013	0.0026	−39.379	39,671.65	<.0001
List-wide PC[a]	−0.0086	0.0104	−0.825	97.51	.4112
Item-specific PC[a]	−0.0135	0.0074	−1.827	435.59	.0684
Contingency[a]	−0.0207	0.0059	−3.497	48,890.40	.0005
ResponseProbability[a]	−0.1234	0.0133	−9.256	49,487.44	<.0001
Congruency:list-wide PC[a]	0.0651	0.0081	8.071	101.81	<.0001
Congruency:item-specific PC[a]	0.0162	0.0102	1.590	939.28	.1121
List-wide PC[a]:item-specific PC[a]	−0.0123	0.0109	−1.132	97.28	.2606
Congruency:list-wide PC [a]:item-specific PC[a]	0.0140	0.0120	1.167	94.08	.2461

Note. *b* = coefficient estimates; *df* = degrees of freedom computed with the Satterthwaite's approximation; PC = proportion congruency; iRTpreceding = inverse-transformed response times of the preceding trial.
[a] These effects refer to trial-level variables as explained in Assessing the Magnitude of List-Wide and Item-Specific PC Effects section.

effect was significantly modulated by trial-level, list-wide PC ($p < .0001$, $d_r = 1.27$, $d_S = 0.80$, dominance = 92.86%), showing that as trial-level, list-wide PC increased, the Stroop effect increased, revealing the effect of proactive control. By contrast, the linear mixed model analysis did not reveal a significant modulation of the Stroop effect by trial-level, item-specific PC ($p = .1121$, $d_r = 0.62$, $d_S = 0.05$, dominance = 70.41%). Moreover, the three-way interaction between the Stroop effect, trial-level, list-wide PC and trial-level, item-specific PC was not significant ($p = .2461$, $d_r = 0.16$, $d_S = 0.12$, dominance = 57.14%), suggesting that trial-level, list-wide PC and trial-level, item-specific PC did not interact in modulating the Stroop effect. Lastly, our analysis revealed a significant effect of trial-level contingency ($p = .0005$), indicating that participants responded faster when trial-level contingency was higher.

To verify whether the effect of the congruency:trial-level, item-specific PC interaction was hindered by the three-way interaction, we also performed the continuous full two-way model, excluding the three-way interaction and keeping the two double interactions (congruency:trial-level, list-wide PC and congruency:trial-level, item-specific PC; see Table 1). This analysis confirmed the results reported above, and the interaction between trial-level, item-specific PC and congruency remained nonsignificant ($p = .0920$, $d_r = 0.50$, $d_S = 0.06$, dominance = 72.45%), confirming that in the peripheral task, we did not find a significant modulation of the Stroop effect by trial-level, item-specific PC. Lastly, the effect of trial-level contingency was again significant ($p = .0004$), showing that participants responded faster when trial-level contingency was higher (see Table S1.7 in the additional online materials at https://osf.io/aeh2d).

To confirm these results, random coefficient analysis was then performed on the iRTs using the continuous full model, namely the model including the three-way interaction. All the effects of the confounding predictors on participants' iRTs found in both linear mixed model analyses were confirmed (all $ps < .004$, see Table S1.8 in the additional online materials at https://osf.io/aeh2d), except for HorizontalStimulus. Random coefficient analysis results regarding our predictors of interest partially replicated linear mixed model results. Indeed, we similarly found a significant Stroop effect ($p < .0001$, $d_r = 2.32$, dominance 98.98%), that is, longer iRTs for incongruent trials, and a significant modulation of the Stroop effect by trial-level, list-wide PC ($p < .0001$, $d_r = 0.86$, dominance = 77.55%). Moreover, the interaction between congruency and trial-level, item-specific PC was still not significant ($p = .9344$, $d_r = 0.01$, dominance = 46.94%), replicating previous analysis that failed to reveal the effect of reactive control in modulating the Stroop effect. However, the results of the three-way interaction were in contrast with the linear mixed model ones. Indeed, we found that trial-level, list-wide PC and trial-level, item-specific PC interacted significantly in modulating the Stroop effect ($p = .0009$, $d_r = 0.35$, dominance = 65.31%). Lastly, the effect of trial-level contingency was significant ($p = .0001$, $d_r = -0.40$), confirming previous results.

Since we found that the effect of trial-level contingency was always significant while the effect of reactive control was never significant, we hypothesized that we did not find it because trial-level contingency might have explained all the variance that could have been explained by the reactive control modulation of the Stroop effect. We thus performed a control analysis, running again the continuous full model after excluding trial-level contingency (continuous full NoContingency model; see Table 1) both using linear mixed model and random coefficient analysis (see Tables S1.9–S.10 in the additional online materials at https://osf.io/aeh2d). These analyses confirmed the results of the previous ones, except for the fact that, by removing trial-level contingency, the interaction between congruency and trial-level, item-specific PC became significant (both $ps < .0001$, $d_s = 0.84$ and $d_r = 0.83$, respectively). Of note, the inclusion of trial-level contingency in the continuous full model was justified and improved the model fit, $\chi^2(1) = 30.1$, $p < .0001$.

Internal Reliability of List-Wide and Item-Specific PC Effects

We assessed the internal reliability of our effects of interest using a linear mixed model analysis (continuous full model) to explain

intrasubject/intertrial variance, with the aim of obtaining more precise estimates of it.

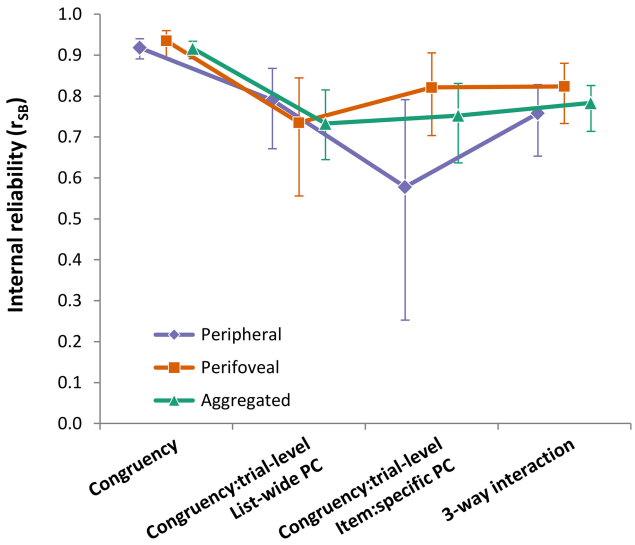
As expected, the internal reliability estimate of the Stroop effect was the highest and least variable among our effects of interest, with a median r_{SB} value of .92 and a 95% CI of [.89, .94]. The internal reliability of the proactive control effect had a median r_{SB} value of .79 and a 95% CI of [.67, .87], while the median r_{SB} internal reliability of the reactive control effect was .58 with a 95% CI of [.25, .79]. Finally, the internal reliability of the three-way interaction was similar to that of proactive control, with a median r_{SB} value of .76 and a 95% CI of [.65, .83] (see Figure 5 and Figure S1.9 in the additional online materials at https://osf.io/aeh2d).

Discussion

The results of our previous study (Viviani et al., 2024) showed that the peripheral spatial Stroop is an experimental paradigm suitable for yielding a complete Stroop interference effect whose magnitude is also large and robust to analytic flexibility with adequate and robust internal reliability. In contrast to our previous study (Viviani et al., 2024), in which we just assessed the magnitude of the Stroop effect, here we used the same experimental paradigm but with different manipulations with the aim of measuring, simultaneously, the effect of proactive and reactive control in modulating the Stroop effect, as well as their interaction, while controlling for low-level effects, including contingency.

Figure 5

Internal Reliability Results



Note. Internal reliability (r_{SB}) of the experimental effects of interest (congruency, Stroop effect; congruency:trial-level, list-wide PC, proactive modulation of Stroop effects; congruency:trial-level, item-specific PC, reactive control modulation of Stroop effects; congruency:trial-level, item-specific PC:trial-level, list-wide PC, interaction between reactive and proactive control in modulating the Stroop effects) in the three analyses. Error bars represent the nonparametric 95% confidence intervals. PC = proportion congruency. See the online article for the color version of this figure.

The analysis assessing the magnitude of such effects revealed different results based on the analytical approach employed. As explained in the methods, the standard general linear model-based analysis is largely incompatible with our experimental design and was thus only used to assess overall Stroop effects. The two multilevel analytical approaches, which were instead more adequate for our purposes, showed a partially contrasting scenario. Indeed, both approaches converged on the existence of a proactive control mechanism modulating the Stroop effect, which was shown to be a large and universal effect, and on the absence of a reactive control mechanism that independently modulated the Stroop effect. On the contrary, a significant interaction of proactive and reactive control in modulating the Stroop effect was found only in the random coefficient analysis-based analysis but not in the linear mixed model-based one, casting shadows on the robustness of such an effect.

A further consistent aspect revealed by our analyses is the relation between trial-level, item-specific, PC-induced reactive control and trial-level contingency. Indeed, they interestingly revealed that when trial-level contingency was statistically controlled for, the trial-level, item-specific, PC-dependent modulation of the Stroop effect was not significant, thus confirming the strong influence of contingency on item-specific, PC-induced reactive control, as suggested by Schmidt (2019) (see also Schmidt & Besner, 2008). Moreover, the fact that the effect of reactive control was unveiled after removing trial-level contingency from the model provides support to our methodological and analytical approaches, which allowed us to control for the impact of contingency at the statistical level, while this is practically impossible to do in a purely methodological way (i.e., with the design-level control). This point will be addressed more in more detail in the general discussion.

Lastly, the internal reliability of the two significant effects of interest was quite high. Specifically, the Stroop effect had a very high internal reliability, characterized also by little variability, whereas the proactive control reliability was more variable but still quite good (see Figure 5).

The inconsistencies regarding the interplay between proactive and reactive control observed in the present experiment could be in part explained by the peripheral spatial Stroop weaknesses assumed also in our previous study (Viviani et al., 2024) and related to the peripheral visual presentation of the stimuli. Specifically, the peripheral arrangement of the stimuli promotes the employment of visuospatial attentional shifts and eye movements to fixate the stimulus and better perceive it, which is a necessary processing step to retrieve the PC specifically associated with the item and then employ reactive control accordingly. However, these processing steps probably delay the employment of reactive control, as compared to a task using a perifoveal arrangement of the stimuli. This might have hindered the strength of reactive control and, consequently, its interaction with proactive control, thus not allowing us to detect it consistently using different analytical approaches.

Therefore, the use of the perifoveal task in the second experiment could help us shedding light on the inconsistencies in our results, since we previously found that this experimental paradigm not only overcomes the weaknesses of the peripheral task, but it is also the best alternative to it among all the task versions considered (Viviani et al., 2024). Indeed, by presenting the stimuli in the perifoveal vision, the perifoveal spatial Stroop task does not require visuospatial attentional shifts or eye movements, thus it may favor a faster and more efficient reactive control employment.

Experiment 2—Perifoveal

Method

We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, all manipulations, and all measures in the study. All inclusion/exclusion criteria were established prior to data analysis. All data and materials, as well as the code used to run the experimental tasks and generate and analyze the data of the current study, are available in our project repository on the OSF platform at https://osf.io/qmu7g. No part of the study, including the analyses, was preregistered.

Procedure and Experimental Tasks

In this experiment, we used a perifoveal spatial Stroop task (Viviani et al., 2024), wherein the target arrow could appear in one of four centrally located spatial positions so that both the task-relevant and task-irrelevant pieces of information could be seen in perifoveal vision. To do so, a different fixation screen was displayed, consisting of a vertically oriented thin black cross (30×30 pixels) enclosed in the partial outline of a black square (94×94 pixels) presented at the center of the screen. The partial outline of the square around the fixation cross created the impression of four small squares, allowing us to manipulate the position inside the fixation stimulus. Therefore, the target arrow was presented within one of these apparent small squares, and participants were required to indicate its pointing directions regardless of its position. Apart from this aspect, the experimental task and procedures were the same as in Experiment 1.

As for the peripheral experiment, we conducted control analyses on block-level, item-specific PC to assess it worked in the expected direction in the two list-wide PC blocks, separately, and this was the case, $ts(77) = 2.01$ and 4.46, $ps < .05$ and .001, $ds = 0.23$ and 0.50.

Data Analysis

The analyses were performed as in Experiment 1. As in Experiment 1, we excluded from the analyses training trials, habituation trials, and buffer trials at the beginning of each subblock. From the resulting 49,920 experimental trials, we also excluded trials with incorrect responses ($n = 1{,}885$, corresponding to 3.78% of the experimental trials), missed responses ($n = 57$, corresponding to 0.11% of the experimental trials), and RTs shorter than 150 ms ($n = 0$), which were all treated as errors, and post-error trials ($n = 1{,}667$, corresponding to 3.34% of the experimental trials).

We checked for the presence of participants with low compliance, defined as those having either a mean iRTs more than three standard deviations away from the sample mean or a mean accuracy lower than 70% (i.e., the level used in the practice block). Based on these criteria, no participant was excluded from the analyses (see Figure S2.2 in the additional online materials at https://osf.io/aeh2d).

Participants

For this experiment, we recruited 78 participants (41 female and 37 male participants; $M_{age} = 24.21$ years, $SD = 6.43$ years). Participants reported demographic information by completing free-response boxes for age gender (the labels "male," "female," or "other/prefer to not respond" were provided). No other demographic information was collected. The sample comprised four left-handed participants (EHI scores < -50) and 13 ambidextrous participants (EHI scores between -50 and 50). No participants reported suffering from neurological or psychiatric disorders or being under medication. Participants gave their informed consent to participate in the study, which was conducted in accordance with the ethical standards of the 2013 Declaration of Helsinki for human studies of the World Medical Association. The study was approved by the Ethical Committee for the Psychological Research of the University of Padova.

Participants consisted of a convenience sample recruited using researchers' personal networks and were not compensated for their participation. A power analysis was performed as in Experiment 1; again, we decided to recruit as many participants as possible, exceeding the required sample size, to be able to detect effects even smaller than expected (by increasing the statistical power of our analyses) and to increase the precision of the experimental effects estimates. It is important here to note that linear mixed models tend to provide higher power than standard general linear model approaches like the one-sample *t* test we used here.

Results

Magnitude of List-Wide and Item-Specific PC Effects

As for the analyses on the peripheral task, here, we report only the results on iRTs, since the distribution of RTs was heavily right-skewed and the residuals of the analyses on both RTs and lnRTs violated the assumption of normality (see Figures S2.1 and S2.3–S2.5 in the additional online materials at https://osf.io/aeh2d).

General linear model-based analyses using *t* tests were first used to explore the overall Stroop effect (i.e., aggregating across list-wide PC and item-specific PC values). These analyses revealed that participants responded significantly slower to incongruent as compared to congruent trials, $t(77) = 33.43$, $p < .0001$, with a very large effect size ($d = 3.79$) and 100% dominance (see Table S2.1 in the additional online materials at https://osf.io/aeh2d).

Then, we performed linear mixed model analyses using both the continuous and the discrete full models. We found that the former provided a better fit to the data (AIC = 29,422 vs. 29,511 of the full discrete model). Then, we compared the continuous full model with the continuous reduced model (see Table 1) performing the log-likelihood ratio test, which revealed that the continuous full model was justified, $\chi^2(7) = 2{,}971.6$, $p < .0001$, suggesting that the inclusion of confounders increased the model fit. As such, here we report the results of the analysis performed on iRTs using the full model with continuous variables.

The conditional R^2 of the linear mixed model was 0.63% and 0.84% of the observations was removed as outliers (>3 absolute standardized residuals) to mitigate the stress of the model fit (i.e., to improve the normality of the residuals, see Figure S2.3 in the additional online materials at https://osf.io/aeh2d). All the confounding predictors of our continuous full model significantly modulated participants' iRTs (all $ps < .04$, see Table 3). We found that participants were faster as trials went on and when they responded to stimuli appearing in the lower and right halves of the screen using the middle finger and the right hand. Moreover, there was a significant temporal dependency in iRTs (i.e., a positive correlation between iRTs in the current and preceding trial), and the responses were faster when the probability of the response (trial-level ResponseProbability) was higher.

Table 3
Results of the Linear Mixed Model Analysis for Experiment 2 (Continuous Full Model)

Effect	*b*	*SE*	*t*	*df*	*p*
Intercept	−2.4845	0.0448	−55.449	128.21	<.0001
Trial	−0.1018	0.0034	−29.687	2,686.32	<.0001
Congruency	0.3729	0.0177	21.059	138.27	<.0001
iRTpreceding	0.0586	0.0017	35.077	45,622.01	<.0001
HorizontalStimulus	−0.0080	0.0039	−2.072	4,301.02	.0383
VerticalStimulus	0.0110	0.0038	2.888	3,793.44	.0039
HorizontalResponse	−0.0608	0.0038	−16.143	31,342.78	<.0001
VerticalResponse	−0.0723	0.0037	−19.484	29,671.40	<.0001
List-wide PC[a]	−0.0002	0.0145	−0.012	78.89	.9907
Item-specific PC[a]	−0.0025	0.0112	−0.223	262.95	.8238
Contingency[a]	−0.0516	0.0085	−6.062	40,688.90	<.0001
ResponseProbability[a]	−0.1572	0.0189	−8.300	39,539.08	<.0001
Congruency:list-wide PC[a]	0.0668	0.0119	5.606	77.49	<.0001
Congruency:item-specific PC[a]	−0.0066	0.0157	−0.416	422.46	.6774
List-wide PC [a]:item-specific PC[a]	−0.0311	0.0149	−2.085	78.49	.0403
Congruency:list-wide PC[a]:item-specific PC[a]	0.0454	0.0204	2.226	78.29	.0289

Note. *b* = coefficient estimates; *df* = degrees of freedom computed with the Satterthwaite's approximation; PC = proportion congruency; iRTpreceding = inverse-transformed response times of the preceding trial.
[a] These effects refer to trial-level variables as explained in Assessing the Magnitude of List-Wide and Item-Specific PC Effects section.

For what concerns our predictors of interest, the Stroop effect (congruency) significantly modulated participants' iRTs ($p < .0001$), who showed slower responses to incongruent trials, with a very large effect size ($d_r = 3.01$, $d_S = 1.79$) and a dominance value of 100%, which indicates that all participants showed a positive Stroop effect. As for the peripheral task, trial-level, list-wide PC significantly modulated the Stroop effect ($p < .0001$, $d_r = 0.99$, $d_s = 0.64$, dominance = 82.05%), that is, there was an effect of proactive control, with larger Stroop effects as trial-level, list-wide PC increased. By contrast, the Stroop effect was not significantly modulated by trial-level, item-specific PC ($p = .6774$, $d_r = -0.11$, $d_s = -0.02$, dominance = 42.31%). In contrast to the peripheral task, here, the three-way interaction was significant ($p = .0289$, $d_r = 0.31$, $d_s = 0.25$, dominance = 57.69%), revealing that the trial-level, list-wide PC and trial-level, item-specific PC interacted in modulating the Stroop effect, that is, when both were high (namely, both proactive and reactive control were low), the Stroop effect was larger. Lastly, the effect of trial-level contingency was again significant ($p < .0001$), indicating that participants responded faster when trial-level contingency was higher. As opposed to the analysis on the peripheral task, here we do not report the continuous full two-way model, since the three-way interaction was significant (but see Table S2.2 in the additional online materials at https://osf.io/aeh2d).

We then performed a random coefficient analysis on the continuous full model, which basically confirmed the results of the linear mixed model analysis. Indeed, all the confounding predictors were in the same direction as the linear mixed model results, and all but HorizontalStimulus and VerticalStimulus were significant (all $ps < .0005$). We also found a very large Stroop effect ($p < .0001$, $d_r = 2.09$, dominance 98.72%), as well as its significant modulation by trial-level, list-wide PC ($p < .0001$, $d_r = 0.60$, dominance = 71.79%). In line with all previous results, the congruency:trial-level, item-specific PC interaction was not significant ($p = .7178$, $d_r = -0.04$, dominance = 51.28%), showing no reactive control effect on the Stroop effect. The random coefficient analysis also confirmed the three-way interaction found with the linear mixed model analysis on the perifoveal task, showing that trial-level, list-wide PC and trial-level, item-specific PC significantly interacted in modulating the Stroop effect ($p = .0006$, $d_r = 0.41$, dominance = 65.38%). Also, the effect of trial-level contingency was again significant as in all our analyses ($p = .0001$, $d_r = -0.47$; see Table S2.8 in the additional online materials at https://osf.io/aeh2d).

Additionally, we performed the same control analysis described for the peripheral experiment to verify whether the nonsignificant effect of reactive control was due to the effect of trial-level contingency by rerunning the continuous full model using linear mixed model and random coefficient analyses but excluding trial-level contingency (continuous full NoContingency model; see Table 1). The results were confirmed but, again, the interaction between congruency and trial-level, item-specific PC became significant after removing trial-level contingency (both $ps < .0001$, $d_s = 0.71$ and $d_r = 0.74$, respectively). Of note, the inclusion of trial-level contingency in the model was justified and improved the model fit, $\chi^2(1) = 18.0$, $p < .0001$ (see Tables S2.9 and S2.10 in the additional online materials at https://osf.io/aeh2d).

Finally, we conducted an additional control analysis to evaluate the robustness of the three-way interaction of interest. Considering that trial-level, item-specific PC and contingency are correlated, we included a further three-way interaction as a confounding factor in the continuous full model, wherein trial-level, item-specific PC was replaced with trial-level contingency (referred to as continuous full control model; see Table 1). Consequently, this means including the interaction between congruency, trial-level contingency, and trial-level, list-wide PC. It is important to note, however, that the interaction between contingency and congruency is not theoretically well-founded, as contingency, being a low-level learning effect, should not depend on and interact with congruency. There is indeed evidence that a contingency effect can be observed even when congruency cannot be computed, such as, when the stimuli imply two features having an arbitrary relationship (Schmidt & Besner, 2008). This notwithstanding, the results of this control analysis further confirmed our previous

results, that is, the three-way interaction between congruency, trial-level, list-wide and item-specific PCs remained significant, both in linear mixed model and random coefficient analyses (both $ps < .0001$, $d_s = 0.23$ and $d_r = 0.50$, respectively; see Tables S2.11 and S2.12 in the additional online materials at https://osf.io/aeh2d).

Internal Reliability of List-Wide and Item-Specific PC Effects

The internal reliability estimate of the Stroop effect was the highest and least variable among our effects of interest, with a median r_{SB} value of .94 and a 95% CI of [.89, .96]. The internal reliability of proactive control had a median r_{SB} value of .74 and a 95% CI of [.56, .84], while the median r_{SB} internal reliability of reactive control was .82 with a 95% CI of [.70, .91]. Finally, the internal reliability of the three-way interaction was quite high, with a median r_{SB} value of .82 and a 95% CI of [.73, .88] (see Figure 5 and Figure S2.9 in the additional online materials at https://osf.io/aeh2d).

Discussion

As discussed in our previous study (Viviani et al., 2024) and in the peripheral experiment discussion (see the Discussion section), the perifoveal task has methodological advantages over the peripheral one, while also showing good statistical properties, as it ensures a large and reliable Stroop effect. Although, so far, this experimental paradigm was tested when only congruency was manipulated (see Viviani et al., 2024), we expected that its methodological advantages over the peripheral task could be extended over different experimental manipulations, such as those used in the present study. As such, we predicted that we could have obtained more reliable and robust results by using the perifoveal Stroop task to simultaneously measure the effect of proactive and reactive control in the modulation of the Stroop effect, as well as their interaction, while also controlling for contingency and other low-level effects.

The results were in line with our predictions and the pattern and magnitude of our effects of interest were consistent across the two multilevel analytical approaches, which showed a significant modulation of the Stroop effect by proactive control alone and by the interaction between proactive and reactive control. Thus, as compared to the peripheral experiment, using the perifoveal task we obtained evidence not only for the effect of proactive control but also for that of the three-way interaction, which was probably favored by the nature of the task that minimized the potential effect of confounders related to the peripheral allocation of attention. Indeed, by allowing a faster identification of the item, the PC associated with it was probably more effectively activated and reactive control was elicited accordingly. As a consequence, the three-way interaction might have had a larger magnitude, and thus might have been more easily detectable.

Although we found reactive control to interact with proactive control in modulating the Stroop effect, no significant distinct reactive control effect emerged. This finding, obtained using an experimental paradigm that is more likely to favor an item-specific, PC-related reactive control employment, provided further evidence for our hypothesis that, when trial-level contingency was included in the model, there was no residual variance left to be explained by trial-level, item-specific PC. Indeed, after trial-level contingency was removed from the model, the effect of trial-level, item-specific, PC-induced reactive control emerged. This control analysis further supported our claim for the need to statistically control for what cannot be controlled for at the design level.

The results for internal reliability were in line with the peripheral ones. The internal reliability of the Stroop effect was very high and showed little variability. Proactive control still had a good internal reliability, but was somewhat lower and more variable than that of the Stroop effect (see Figure 5).

Overall, the methodological premises favoring the perifoveal spatial Stroop task and the greater consistency of the results across different approaches could indicate that the results obtained using the perifoveal task were more robust and trustable. However, since they are in contrast with those obtained using the peripheral task, we performed a between-experiments analysis to verify whether the patterns of results were actually different between the two experiments.

Between-Experiments Comparisons

Method

We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, all manipulations, and all measures in the study. All inclusion/exclusion criteria were established prior to data analysis. All data and materials, as well as the code used to run the experimental tasks and generate and analyze the data of the current study, are available in our project repository on the OSF platform at https://osf.io/qmu7g. The analyses were not preregistered.

Data Analysis

Between-Experiment Differences in List-Wide and Item-Specific PC Effects. We run all the previous analyses but now to compare whether the experimental effects differed among the peripheral and perifoveal experimental tasks. We again checked for the presence of participants with low compliance, defined as those having either a mean iRTs more than three standard deviations away from the sample mean or a mean accuracy lower than 70% (i.e., the level used in the practice block). Based on these criteria, no participant was excluded from the analyses (see Figure S3.2 in the additional online materials at https://osf.io/aeh2d).

First, a general linear model analysis was performed to verify whether the Stroop effects obtained in the two experiments differed. Specifically, we compared the overall Stroop effects using a two-sample t test and computing Cohen's d to obtain the corresponding effect size estimate.

For the linear mixed model analysis, we added to the continuous full model used in the previous analyses the categorical variable experiment, whose value was set to $-.5$ for the peripheral Experiment and .5 for the perifoveal one. As explained above, based on our theoretical assumptions, we a priori decided to test the model including the trial-level estimates of our predictors and to include all plausible confounders. The experiment factor was tested in interaction with those predictors that we expected to be modulated by it. As regards low-level predictors, we hypothesized that experiment, due to the spatial arrangement of the stimuli, interacted with HorizontalStimulus and VerticalStimulus (but not with HorizontalResponse and VerticalResponse since the

response effectors were the same), and with trial-level contingency since the difference in the stimuli could have affected the contingency effects. Moreover, since we were interested in whether the proactive and reactive control interaction in the Stroop effect modulation differed among the two experiments, the experiment factor was also tested in interaction with them in the fixed part, thus including in the model a four-way interaction. The formula for the final model, referred to as the continuous full Between4-way model, is: RT ~ iRTpreceding + Trial + HorizontalResponse + VerticalResponse + Trial-Level ResponseProbability + Experiment × (HorizontalStimulus + VerticalStimulus + Trial-Level Contingency + Trial-Level, List-Wide PC × Trial-Level, Item-Specific PC × Congruency) + (Trial-Level, List-Wide PC × Trial-Level, Item-Specific PC × Congruency|Subject).

We inspected the results to identify whether the effect of the predictors we tested in interaction with the experiment factor was significant or not and, in the latter case, we refitted the model without such interactions to exclude the possibility that they could have affected the estimation of the other effects. It is important here to note that the results were essentially the same. Then, we used the random slopes for each participant to obtain the participant's mean for each combination of experimental conditions and compared the experimental effects between the two experiments using independent-sample *t* tests.

For the random coefficient analysis, we ran linear regressions at the subject level using the same regression model as the within-subjects analysis (continuous full model), that is, using the following formula: RT ~ iRTpreceding + Trial + HorizontalStimulus + Vertical Stimulus + HorizontalResponse + VerticalResponse + Trial-Level Contingency + Trial-Level ResponseProbability + Trial-Level, List-Wide PC × Trial-Level, Item-Specific PC × Congruency.

After excluding outliers exceeding 3 *SD*s, we refitted the model and we compared whether the parameters of interest differed between the two experiments using independent-sample *t* tests.

These analyses could also help us in case of inconsistencies in the results of the two within-subjects analyses as, by being performed on the two data sets together, it would provide an overall result based on a larger sample.

Assessing the Internal Reliability of List-Wide and Item-Specific PC Effects. The internal consistency of the experimental effects of interest was assessed for the linear mixed model results in the aggregate sample as described in Experiment 1.

Results

Between-Experiments Differences in List-Wide and Item-Specific PC Effects

We report here only the results of the analyses performed on iRTs for the same reasons explained above. The general linear model-based analysis using a two-tailed independent-samples *t* test showed that the mean Stroop effects in the two experiments differed significantly, with the perifoveal one yielding the larger Stroop effect, $M_{diff} = 0.0832$, $t(174) = 4.77$, $p < .0001$, $d = 0.72$.

Regarding linear mixed model analysis, the continuous full Between4-way model revealed that the low-level confounders that we tested in interaction with experiment were significant, confirming our assumption that the experiments differed for the horizontal position of the stimulus ($p = .0172$) and for the vertical position of the stimulus ($p < .0001$). Moreover, the confounding predictors tested alone were all significant ($p < .0001$, see Table S3.1 in the additional online materials at https://osf.io/aeh2d). Specifically, participants responded faster as trials went on, responded faster to stimuli appearing in the upper and right positions, using the middle finger and the right hand, and when trial-level ResponseProbability was higher. We also found a significant temporal dependency in iRTs.

For what concerns the predictors of interest tested in interaction with the experiment factor, the four-way interaction was not significant ($p = .1588$), revealing that the interaction between trial-level, list-wide PC and trial-level, item-specific PC in modulating the Stroop effect was not different between the two experiments. The Stroop effect and the effects of proactive and reactive control did not differ between the experiments ($p = .4792$, $p = .6713$, and $p = .2110$, respectively). By contrast, the effect of trial-level contingency was significantly different between the two experiments ($p = .0002$), with a greater effect of trial-level contingency in the perifoveal experiment as compared to the peripheral experiment. Regarding the predictors of interest not tested in interaction with the experiment, the results confirmed a significant Stroop effect (congruency), with a very large effect size ($p < .0001$, $d_r = 3.33$, $d_S = 1.95$) and a dominance of 100%, indicating that all participants responded slower to incongruent trials. Moreover, trial-level, list-wide PC significantly modulated the Stroop effect ($p < .0001$, $d_r = 1.15$, $d_s = 0.71$, dominance 86.36%), revealing a significant effect of proactive control. Similar to previous analyses, we did not find a significant trial-level, item-specific PC modulation of the Stroop effect ($p = .5849$, $d_r = 0.12$, $d_s = 0.02$, dominance 55.68%). Interestingly, the three-way interaction between trial-level, list-wide PC, trial-level, item-specific PC and congruency was significant ($p = .0094$, $d_r = 0.25$, $d_s = 0.20$, dominance 59.09%), suggesting that trial-level, list-wide PC and trial-level, item-specific PC interacted in modulating the Stroop effect. This result was in line with the perifoveal task results but not with the peripheral ones, and provided additional evidence for the interaction of proactive and reactive control in the modulation of the Stroop effect. Lastly, the effect of trial-level contingency was again significant ($p < .0001$), indicating that participants responded faster when trial-level contingency was higher (see Table S3.1 in the additional online materials at https://osf.io/aeh2d).

As explained in the Method section, we performed the same model but after excluding the interactions between experiment and the predictors that resulted nonsignificant in the previous analysis, which essentially consisted in removing the interactions between experiment and the experimental effects of interest (i.e., the effect of proactive and reactive control, as well as their interaction, in modulating the Stroop effect). This model (continuous full Between model) is basically identical to the ones performed on the two experiments separately, but here it was run on the two data sets aggregated together. As such, since the separate within-subjects analyses on the two experiments revealed contrasting results, this analysis also helped us to resolve the inconsistencies between them, by verifying whether, by removing experiment, the three-way interaction survived.

This analysis confirmed the results reported above for the continuous full Between4-way model. The conditional R^2 of the model was 0.72% and 0.95% of the observations was removed as outliers (>3 absolute standardized residuals) to mitigate the stress of the

model fit (i.e., to improve the normality of the residuals, see Figure S3.3 in the additional online materials at https://osf.io/aeh2d). The effects of low-level confounders, as well as their interaction with the experiment factor, remained significant and in the same direction as in the previous analysis (see Table 4). Regarding the predictors of interest, the Stroop effect was again significant with a large effect size ($p < .0001$, $d_r = 3.28$, $d_s = 1.95$) and complete dominance. Trial-level, list-wide PC still significantly modulated the Stroop effect ($p < .0001$, $d_r = 1.16$, $d_s = 0.72$, dominance = 87.5%), confirming the effect of proactive control. Similarly to the previous analysis, the trial-level, item-specific PC modulation of the Stroop effect was not significant ($p = .4792$, $d_r = 0.15$, $d_s = .02$, dominance = 56.82%). Lastly, the three-way interaction between trial-level, list-wide PC, trial-level, item-specific PC and congruency was again significant ($p = .0150$, $d_r = 0.24$, $d_s = 0.19$, dominance = 55.68%), indicating that trial-level, list-wide PC and trial-level, item-specific PC interacted in modulating the Stroop effect, and when they were both high, participants showed larger Stroop effects. Thus, these results confirmed both those from the perifoveal experiment and those reported above for the between-experiments continuous full Between4-way model.

Random coefficient analysis was then performed to confirm between-experiments linear mixed model results, and this was the case. Indeed, all the predictors of interest were not significantly different in the two experiments (all ps $> .14$; see Table S3.3 in the additional online materials at https://osf.io/aeh2d). Therefore, this analysis confirmed the robustness of the results obtained with random coefficient analysis. Confounding predictors were all significant (all ps $< .04$; see Table S3.3 in the additional online materials at https://osf.io/aeh2d) and in the same direction as the previous results, except for VerticalStimulus, for which we obtained contrasting results in the two experiments. Here, we found that participants were faster in responding to stimuli appearing in the upper part of the screen, as for the peripheral experiment. In line with previous results, we found a very large Stroop effect ($p < .0001$, $d_r = 2.18$, dominance = 98.86%), which was significantly modulated by trial-level, list-wide PC ($p < .0001$, $d_r = 0.71$, dominance = 75%) but not by trial-level, item-specific PC ($p = .8147$, $d_r = -0.02$, dominance = 48.86%). Moreover, trial-level, list-wide PC and trial-level, item-specific PC interacted significantly in modulating the Stroop effect ($p < .0001$, $d_r = 0.37$, dominance = 65.34%), which is consistent with the random coefficient analysis results from both experiments. The effect of trial-level contingency was also significant ($p < .0001$, $d_r = -0.43$; see Table S3.3 in the additional online materials at https://osf.io/aeh2d).

Given that between-experiments results also showed that trial-level contingency was significant but the congruency by trial-level, item-specific PC interaction was not, we further tested our hypothesis that we did not find it because trial-level contingency explained all the variance that could have been explained by the reactive control modulation of the Stroop effect. For both the linear mixed model and the random coefficient analysis analyses, the resulting continuous full BetweenNoContingency model showed the same pattern of results, both for confounders and predictors of interest, with the only exception that, after removing trial-level contingency from the model, the interaction between congruency and trial-level, item-specific PC became significant (ps $< .0001$) and with large effect sizes ($d_r = 0.72$ and 0.76, respectively; see Tables S3.5 and S3.6 in the additional online materials at https://osf.io/aeh2d). Of note, the inclusion of trial-level contingency in the model was justified and improved the model fit, $\chi^2(2) = 138.4$, $p < .0001$.

Table 4
Results of the Linear Mixed Model Analysis for the Between-Studies Analysis (Continuous Full Between Model)

Effect	*b*	*SE*	*t*	*df*	*p*
Intercept	−2.2029	0.0278	−79.225	273.92	<.0001
Trial	−0.0967	0.0020	−48.649	5,466.32	<.0001
Congruency	0.3612	0.0105	34.470	311.05	<.0001
iRTpreceding	0.0568	0.0010	58.150	102,923.55	<.0001
HorizontalStimulus	−0.0103	0.0023	−4.394	17,358.21	<.0001
VerticalStimulus	−0.0330	0.0023	−14.442	20,631.10	<.0001
HorizontalResponse	−0.0463	0.0022	−21.001	72,402.04	<.0001
VerticalResponse	−0.0899	0.0022	−41.542	74,362.37	<.0001
List-wide PC[a]	−0.0057	0.0087	−0.653	175.99	.5144
Item-specific PC[a]	−0.0083	0.0064	−1.291	656.11	.1972
Contingency[a]	−0.0356	0.0050	−7.152	91,065.10	<.0001
ResponseProbability[a]	−0.1335	0.0111	−11.977	92,613.89	<.0001
Congruency:list-wide PC[a]	0.0658	0.0069	9.590	177.58	<.0001
Congruency:item-specific PC[a]	0.0063	0.0089	0.708	1,198.54	.4792
List-wide PC[a]:item-specific PC[a]	−0.0214	0.0090	−2.389	175.94	.0180
Congruency:list-wide PC [a]:item-specific PC[a]	0.0279	0.0113	2.458	174.02	.0150
Experiment	−0.4712	0.0448	−10.523	182.63	<.0001
HorizontalStimulus:Experiment	−0.0098	0.0041	−2.376	17,548.36	.0175
VerticalStimulus:Experiment	0.1047	0.0041	25.703	22,145.14	<.0001
Contingency[a]:Experiment	−0.0229	0.0053	−4.357	336.10	<.0001

Note. *b* = coefficient estimates; *df* = degrees of freedom computed with the Satterthwaite's approximation; PC = proportion congruency; iRTpreceding = inverse-transformed response times of the preceding trial.
[a] These effects refer to trial-level variables as explained in Assessing the Magnitude of List-Wide and Item-Specific PC Effects section.

Lastly, as in the perifoveal experiment, we conducted an additional control analysis to evaluate the robustness of the significant three-way interaction of interest, adding, as confounding factor, the interaction between congruency, trial-level contingency, and trial-level, list-wide PC (continuous full control model). Although this interaction is theoretically unjustified, as explained above, the results of this control analysis further confirmed our previous results, still showing that the three-way interaction between congruency and trial-level, list-wide and item-specific PCs was significant, both in linear mixed model and random coefficient analyses (both $ps < .0001$, $d_s = 0.16$ and $d_r = 0.40$, respectively; see Tables S3.7 and S3.8 in the additional online materials at https://osf.io/aeh2d).

Comparison With the Hypothesized Models

Lastly, we compared our results to the models we put forward in the Introduction to verify which one better explained the patterns we obtained. We decided to compute such a comparison directly on the data aggregated over the two experiments, that is, those used in the between-experiments analysis with the continuous full model. Specifically, we correlated the overall observed pattern of Stroop effects predicted by the linear mixed model analysis (Figure 6A) with those predicted by each of our a priori models to identify which had the highest correlation. We found that the observed Stroop effects were correlated the most with the model assuming an antagonistic interaction with a higher effect of proactive compared to reactive control ($r = .97$; see Figure 6B for all correlations). Moreover, it should be noted that the three additive models cannot be applied to our results, because the three-way interaction was significant; similarly, the models assuming that reactive control has an equal or greater effect compared to the proactive control cannot be applied to our results, because the proactive control effect ($M = 0.132$, $SD = 0.114$) was significantly larger than the (nonsignificant) reactive control one ($M = 0.013$, $SD = 0.084$), $t(175) = 11.20$, $p < .0001$, $d = 0.84$. Therefore, there were only two models that can be applied to our results, that is, those assuming a greater effect of proactive control and its interaction, either antagonistic or synergistic, with reactive control. To confirm that the former better explained our results, we extracted the by-participants patterns of Stroop effects in the four conditions, and correlated them with the patterns of Stroop effects hypothesized by the two models; a paired-sample t test confirmed that the antagonistic model better explained the participants' performance, $t(175) = 3.69$, $p = .0003$, $d = 0.28$.

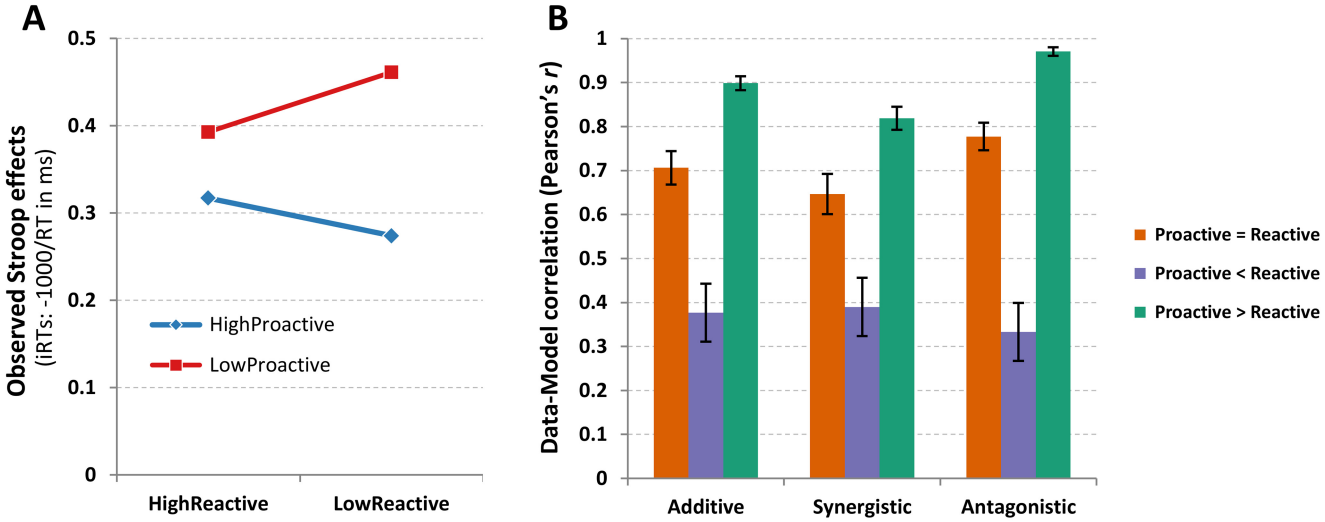
Internal Reliability of List-Wide and Item-Specific PC Effects in the Aggregated Sample

As expected, the internal reliability estimate of the Stroop effect was the highest among our effects of interest, with a median r_{SB} value of .92 and a 95% CI of [.89, .93]. The internal reliability of proactive control had a median r_{SB} value of .73 and a 95% CI of [.64, .82], while the median r_{SB} of reactive control was .75, with a 95% CI of [.64, .83]. Finally, the internal reliability of the three-way interaction was similar, with a median r_{SB} value of .78 and a 95% CI of [.71, .83] (see Figure 5).

General Discussion

According to the DMC model (Braver, 2012; Braver et al., 2007), cognitive control operates via two distinct mechanisms, proactive and reactive control, which are qualitatively different in terms of

Figure 6
Observed Stroop Effects and Data-Model Correlations



Note. (A) The plot shows the observed pattern of Stroop effects, as estimated by the continuous full between model in the linear mixed model analysis, as a function of the level of proactive control (HighProactive, blue [light gray] line, corresponding to low trial-level, list-wide PC conditions; LowProactive, red [dark gray] line, corresponding to high, trial-level, list-wide PC conditions) and reactive control (HighReactive, corresponding to low trial-level, item-specific PC conditions; LowReactive, corresponding to high trial-level, item-specific PC conditions). (B) Correlations between the observed pattern of Stroop effects and the hypothesized models (see the Hypotheses section). Error bars represent the standard error of the correlation. iRTs = inverse-transformed response times; RT = response times. See the online article for the color version of this figure.

their temporal dynamics. Although this model could account for the great variability intrinsic to this ability, the evidence currently available for it is not compelling. In the laboratory, the DMC model has been frequently assessed with the Stroop task, which allows varying the PC at the list-wide and/or at the item-specific levels to specifically target proactive and reactive control, respectively. However, these manipulations have been called into question, especially by the contingency hypothesis, which claims that they elicit low-level processes instead of control-driven ones (Schmidt, 2019; Schmidt & Besner, 2008; Schmidt et al., 2007). Although several confound-minimized manipulations have been proposed to solve this issue, in our view, they still suffer from some limitations and are impractical. Moreover, the two control mechanisms have always been explored separately by implementing each PC manipulation one at a time, which prevents from assessing their specific effects when both are manipulated and, especially, the interplay between the two control mode mechanisms.

Therefore, to date, there is no compelling evidence clearly supporting the existence of two distinct mechanisms, while also controlling for the potential influence of low-level confounders. Our aim here was indeed to tackle this issue by combining multilevel modeling, the state-of-the-art trial-level analytical approach to estimate the Stroop effect and its control-related modulations effectively and reliably (see Viviani et al., 2024), with a novel methodological approach allowing to manipulate both list-wide PC and item-specific PC in a fine-grained way at the trial-by-trial level.

In brief, our main results consistently indicate that proactive control induced by trial-level, list-wide PC manipulations modulated the Stroop effect, whereas trial-level, item-specific, PC-induced reactive control did not, probably due to the confounding effect of contingency. However, both control modes interacted in modulating the Stroop effect.

Advantages of Our Methodological and Analytical Approach

Before elaborating on the findings, it is worthy to discuss the methodological and analytical novelties of the present study, to fully understand their advantages compared to classical approaches primarily implying design-level control only. To make a step further, we used an analysis-level control, leveraging trial-level multilevel modeling to put forward a new approach in which list-wide PC and item-specific PC were manipulated simultaneously. This is indeed a more effective way to (a) directly explore whether proactive and reactive control can coexist, that is, whether each mechanism can be active while also the other is activated; and (b) investigate whether proactive and reactive control interact, since if we assume that they are distinct mechanisms, it is also plausible that they interact as (implicitly) postulated by the DMC model. Moreover, to control for contingency-related effects, we did not balance stimulus-response combinations to make the contingency orthogonal to list-wide PC and item-specific PC as much as possible, and then we controlled for its effect at the statistical level.

Overall, our approach introduced some main novelties, which our data suggested to be advantageous as compared to the traditional approaches used in the literature. The first original aspect of our approach consisted of calculating the trial-wise probabilities of our variables of interest. These provided more realistic estimates of our variables at each trial because they were based on the updating of the trial-by-trial probability based on the trial history. Our expectation of a better model fit when using continuous as compared to discrete values was indeed confirmed by the model comparisons that we performed. Notably, these findings offer indirect support for frameworks that started conceptualizing cognitive control in terms of Bayesian inference (Jiang et al., 2014; Parr et al., 2023), suggesting a potential alignment between the observed results and the principles of the predictive brain (Clark, 2013).

The second novel aspect of our study concerned the analytical approach, as we used multilevel, trial-level modeling, which was particularly suitable for our aims. First, it allowed us to assess all our experimental effects of interest at the same time, that is, the Stroop effect, proactive control, reactive control, and their interaction in modulating the Stroop effect, which was our main aim and was not feasible using classical general linear model-based approaches. As detailed in the Method sections, general linear model analyses were used just to assess the overall Stroop effect (and thus allowing its comparison with the existing studies) but they were not suitable for our experimental design and aim. Indeed, the second important advantage of multilevel modeling was that it ensured that the estimates of our effects were partialed out from the effect of lower-level confounders at the trial level, which represent sources of trial-by-trial noise that can affect the estimation of the effects at the subject level. This advantage, in line with the results of our previous study (Viviani et al., 2024), was confirmed by the comparisons of the full model with the reduced one, which showed that, in all our analyses, the former better explained our data. These results suggest that including in the model low-level predictors based on the characteristics of the task at hand ensured that the estimates of our effects of interest were more accurate and precise since intrasubject/intertrial sources of variance, which contribute to measurement error, were explained. Moreover, using this approach, we successfully cleaned our effects of interest from the effect of contingency which represents a great issue when using PC manipulations as, in our view (see The Stroop Task: PC Manipulations and Their Limitations section), they can hardly be properly controlled at the design level only. Indeed, it should be noted that it is practically impossible to deconfound item-specific PC and contingency measures for congruent trials, as they are both computed in the same way (i.e., the ratio between the occurrence of congruent trials and the total number of trials for that item within a block). Our results indeed indicate that controlling for the effects of contingency at the analysis level, that is, by including it in the model, represents a valid alternative to controlling for it at the design level. We indeed found that item-specific, PC-related results changed dramatically when contingency was removed from the statistical model as compared to when it was included. This indicates that not controlling for contingency-related effects severely affected the results, leading to misleading conclusions about spurious effects of reactive control.

List-Wide and Item-Specific PC Effects and Between-Experiments Differences

The peripheral and the perifoveal experiments yielded contrasting results for what concerns the three-way interaction (namely, between the Stroop effect, trial-level, list-wide PC and trial-level, item-specific PC), with the former not finding any interaction between proactive and reactive control in Stroop effect modulation

and the latter showing it instead. To shed light on this discrepancy, a between-experiments analysis was performed, by running a model that compared the effects of interest as well as the low-level ones that were assumed to be different (i.e., stimulus position and contingency) to assess whether they differed between the two experiments. low-level results confirmed our assumption that the two experiments implied different effects of stimulus position, which was quite predictable as the perifoveal task had been intentionally designed to reduce visuospatial attention shifts and/or eye movements. Moreover, the difference in contingency might also be explained in terms of different spatial arrangements, as the greater effect of contingency in the perifoveal task might have been explained by the lower eccentricity of the stimuli. The keypresses used to provide responses were indeed spatially arranged more similarly to the perifoveal stimuli, as they also had a low eccentricity. This might have led to a greater stimulus-response overlap in the perifoveal task[2] which, in turn, might have favored the learning of stronger stimulus-response associations. No experiment-dependent difference was instead observed for the effects of interest.

Given that between-experiments results revealed that the two experiments were not different for what concerns the effects of interest, we aggregated the two samples to assess the results on both experiments together, as this could tell us more about which pattern of results was more consistent, that is, whether the emerging results on the aggregated sample were more compatible with the peripheral or the perifoveal results. Importantly, the analysis on the aggregated sample, as compared to the ones performed for each experiment, was expected to yield more robust evidence by ensuring more power and more precise estimates. Interestingly, the results of the aggregated analysis reflected those obtained in the perifoveal experiment, although that was the experiment with the smaller sample size. Essentially, we found again that proactive control, but not reactive control, modulated the Stroop effect, and that both control modes interacted in modulating the Stroop effect.

Therefore, the three-way interaction results could seem in contradiction with the between-experiments results that did not reveal any significant difference between the two experiments as, after aggregating the two samples, they showed results in line with one experiment but not with the other. To try to explain this inconsistency, we put forward a possible explanation based on the results of our previous study (Viviani et al., 2024), wherein we found that the greater visuospatial attention shifts characterizing the peripheral task led to an underestimation of the Stroop effect magnitude whereas, when they were reduced, such as in the perifoveal task, the Stroop effect was more robust and larger. As such, these attentional shifts might have reduced the magnitude of the three-way interaction in the peripheral task, making it not detectable with the within-subjects analysis, probably because the peripheral experiment was too underpowered to detect it, but revealing such interaction in the aggregated samples analysis, thanks to the greater power. Of note, the direction of the three-way interaction in the peripheral task tested alone, despite being not significant, was consistent with the direction of the three-way interaction in the perifoveal and in the aggregated sample, supporting our hypothesis that the difference among the two experiments was not qualitative but just quantitative due to the effect of visuospatial attention shifts. This notwithstanding, the difference in the magnitude of the two three-way interactions in the two experiments was not large enough to be significant and thus could not be detected by the between-experiments analysis. Therefore, this interpretation would reconcile our apparently contrasting results, supporting both the presence of distinct patterns, as shown by the within-subjects results, and the absence of evidence for a difference among the experiments, as found in the between-experiments results, allowing us to more safely rely on the aggregated sample results to draw our conclusions.

The robustness of the pattern of results provided by the aggregated sample analysis appears to be well substantiated also by the random coefficient analyses which served us to confirm linear mixed model results. Indeed, the random coefficient analyses did not show contradictory results and always revealed that the Stroop effect was modulated by proactive control alone and by the interaction between proactive and reactive control. Therefore, this further supported our interpretation that the lack of a significant three-way interaction in the peripheral experiment alone using linear mixed model analysis was the consequence of the underestimation of such effect due to the factors discussed above. Still, the fact that this effect emerged only under certain conditions and depending on the analytical approach leads us to suggest taking it with some caution and indicates that further investigations are needed before drawing definitive conclusions about it.

A result that was instead always consistent was the absence of the main effect for the reactive control modulation of the Stroop effect. Indeed, none of our analyses showed a significant interaction between item-specific PC and the Stroop effect (trial-level, item-specific PC and congruency). The absence of such interaction was observed also in one of the control analyses performed on the peripheral experiment using the linear mixed model approach, in which we removed from the model the three-way interaction, as it was not significant, to assess whether it interfered with the estimation of reactive control effects. Even after excluding the three-way interaction, the effect of reactive control by itself did not emerge, suggesting that it was not masked by the three-way interaction.

Therefore, to provide a possible explanation for why trial-level, item-specific, PC-induced reactive control alone never modulated the Stroop effect, we performed control analyses on both single and aggregated samples by excluding from the model the trial-level contingency predictor. This was done because there is extensive literature showing that reactive control is specifically confused with contingency when item-specific PC is manipulated (e.g., Schmidt, 2019; Schmidt & Besner, 2008). Moreover, as explained above, it is very difficult to totally decorrelate contingency and item-specific PC at the design level while also keeping this manipulation item-specific. Therefore, we assumed that the effect of trial-level, item-specific PC and congruency interaction might have been nonsignificant since trial-level contingency included in the model explained all its variance. This was exactly the case: In all the control analyses, the removal of trial-level contingency produced the same pattern of results, except for the interaction between trial-level, item-specific PC and congruency, which became significant after this change. Thus, these consistent results confirmed our assumption that trial-level contingency alone can explain all the variance of reactive control modulation and have several implications, as detailed in the following paragraph.

[2] According to Kornblum (e.g., 1992), stimulus–response overlap is fundamental to yield a complete Stroop effect (see also Viviani et al., 2023), thus the stronger such overlap, the greater the Stroop effect.

First, trial-level, item-specific, PC-induced reactive control and trial-level contingency are intrinsically related since the manipulation used to induce such a mechanism is inevitably confused with contingency, especially for congruent items. Second, as confound-minimized designs that have been proposed in the literature are, in our view, unpractical and partially flawed (see The Stroop Task: PC Manipulations and Their Limitations section), the more adequate approach to control for contingency is the analysis-level one, as it effectively estimated the contingency-related confounding effect, allowing us to suggest that our estimation of reactive control effects was not biased by contingency. As discussed above, the model comparisons provided further evidence in favor of analysis-level control, which showed that the observed data were better explained when confounding predictors, among which trial-level contingency, were included in the statistical model. The third implication regards the fact that, although trial-level, item-specific, PC-induced reactive control alone was not significant, its interaction with proactive control to modulate the Stroop effect was significant even when trial-level contingency was in the model, suggesting that our experimental design was effective at yielding a reactive control effect partialed out from contingency, albeit an indirect one. This claim was further supported by another follow-up control analysis to assess the robustness of the three-way interaction, found significant in the perifoveal and the between-experiments analyses. Given that trial-level, item-specific PC and trial-level contingency were correlated, we tested a model wherein, in addition to the three-way interaction of interest, a further three-way interaction was added as confounding factor, in which, instead of trial-level, item-specific PC, trial-level contingency was included in the three-way interaction to assess whether it interacted with trial-level, list-wide PC in modulating the Stroop effect (continuous full control model). Although we were aware that this interaction was not theoretically well-founded (see Schmidt & Besner, 2008), its inclusion in the model did not affect the effect of our three-way interaction of interest, which remained significant, providing further evidence for the significant role of trial-level, item-specific PC in the three-way interaction and also indicating that it was successfully partialed out from the effect of trial-level contingency (see S2.11, S2.12, and S3.7–S3.8 in the additional online materials at https://osf.io/aeh2d, for the perifoveal and between-experiment results, respectively).

A further result robust to analytical flexibility, and thus consistent across all our analyses, was the Stroop effect modulation by proactive control, which is in line with previous findings (e.g., Bugg & Chanani, 2011; Hutchison, 2011). Indeed, the interaction between trial-level, list-wide PC and congruency was not only always significant in our main linear mixed model and random coefficient analyses, but it was also accompanied by d values that were greater than 0.5, which is considered the threshold of medium effect sizes. In the aggregated sample results, proactive control had a medium-to-large effect size ($d = .71$) and a dominance of almost 87%, indicating that our paradigms were effective in producing a proactive control effect in the expected direction in most of the participants (i.e., larger Stroop effects when trial-level, list-wide PC was higher).

It has to be noted also that this evidence for a proactive control modulation of the Stroop effect was obtained while controlling for possible confounding factors that might influence it. Indeed, whether list-wide PC manipulation is effective in inducing a control mechanism operating at the list-level and not just at the item level has been challenged. Essentially, some authors argued that list-wide PC is inevitably confused with item-specific PC since low-PC blocks are composed of low-PC items and high-PC blocks are composed of high-PC items (e.g., Blais & Bunge, 2010; Blais et al., 2007; Bugg et al., 2008). This is an inevitable consequence of list-wide PC manipulation. Indeed, although we tried to orthogonalize list-wide PC and item-specific PC as much as possible, our predictors were still moderately correlated. However, by using multilevel modeling and including both predictors in the model, we were able to control for that confound. Therefore, since the trial-level, list-wide PC by congruency interaction consistently emerged, as opposed to the trial-level, item-specific PC by congruency interaction, which was never significant, we can reasonably suggest that the observed variance was explained by the proactive (but not by the reactive) control predictor. These results further highlighted the advantages of analysis-level control, which can overcome issues that cannot easily be addressed by the design-level control.

The last important result concerned the magnitude of the Stroop effect. Although in this work we were not specifically interested in the Stroop effect per se, it was essential that our paradigms yielded an effect that was robust, and this was the case. The Stroop effect was indeed observed in all our analyses and was characterized by very large effect sizes. Specifically, in the aggregated analysis, it had a d value of 1.95 and all individuals showed a true positive Stroop effect, that is, they all responded more slowly in incongruent compared to congruent trials, as indicated by a dominance of 100%. These results are in line with those from our previous study, in which we found that both the peripheral and the perifoveal tasks were effective in producing large Stroop effects. However, the present results did not confirm the larger Stroop effects we found for the perifoveal as compared to the peripheral task in our previous study (Viviani et al., 2024). Indeed, albeit in the same direction, this between-experiments difference was not significant in the present study.

Overall, this complex pattern of results allowed us to start to answer one of our research questions, that is, whether proactive control exists per se also while reactive control is present and vice versa, and whether they interact as well. For the reasons explained above, such a question was addressed considering the results produced by the aggregated sample analysis because its higher statistical power ensured more precise estimates of the effects of interest, which, inter alia, did not differ among the two experiments.

Interestingly, our results suggest that list-wide PC-induced proactive control independently operated by modulating the magnitude of the Stroop effect, even when reactive control was manipulated, whereas item-specific, PC-induced reactive control did not affect the Stroop effect by itself, but it only interacted with the proactive control in modulating the Stroop effect. As such, this might indicate that, overall, proactive control is stronger than reactive control, but this claim will be better discussed in the Testing of A Priori Formal Model section. However, care must be taken in interpreting this pattern of results. Indeed, there is no clear evidence against reactive control that exists independently of proactive control because of the limitations of the item-specific PC manipulation that have been pointed out in the literature and confirmed in our work. Indeed, what clearly stood out from our work, and especially from the comparison between the main results and the control ones, was that the effect of reactive control was masked by that of contingency to such an extent that, when both were included in the statistical model, contingency alone explained all the variance. As already

pointed out, this is very likely to be the consequence of the imperfect orthogonalization due to the inevitable overlap of contingency and item-specific PC for congruent trials. However, although the correlation between our trial-level, item-specific PC and trial-level contingency predictors was not that high ($r = .364$ in the aggregated sample, corresponding to 13.2% of shared variance), trial-level contingency still preferentially emerged as a significant modulator of participants' performance, suggesting its stronger influence. This claim seems to be in line with previous works showing that the item-specific PC effect was only incidental and exclusively due to contingency learning (e.g., Schmidt, 2019; Schmidt & Besner, 2008). It is worth noting that the conclusions drawn by the contingency learning account have been quite radical, pointing toward a mere low-level associative learning, rather than conflict-related resolution of the interference in the Stroop task. However, our data do not seem to support such an extreme view either, revealing that a more intermediate and balanced position might better fit the available data.

Essentially, on the one hand, our results are consistent with the contingency learning account for what concerns the inevitable influence of contingency learning on performance. contingency prevailed especially over item-specific, PC-induced reactive control which, when considered by itself, did not survive the presence of such low-level learning effect. Therefore, we agree on the need to control for contingency learning and that considerable attention must be paid in interpreting item-specific, PC-induced reactive control effects. However, our results also revealed that the impact of reactive control still remained significant even after partialing out the effect of contingency, that is, when reactive control interacted with proactive control in modulating the Stroop effect. Therefore, it might be that contingency is stronger than reactive control per se, but this does not necessarily exclude a strategic implementation of control, which can also operate in a reactive way but only when moderated by proactive control levels. Such modulation was, in fact, specifically driven by item-specific, PC-induced reactive control as shown by one of our control analyses, as discussed above (also see the additional online materials at https://osf.io/aeh2d), which allowed us to exclude the role of contingency learning in such a higher-level modulation. More importantly, the pattern of the significant interaction between proactive and reactive control in modulating the Stroop effect offers another alternative explanation for the fact that the reactive control modulation of the Stroop effect did not reach the significance level per se. Indeed, as we discuss in detail below, our results indicate that the effect of item-specific, PC-induced reactive control emerged only when the level of proactive control was low. However, in our models the effect of the interaction between congruency and the trial-level, item-specific PC estimate is conditional to an intermediate level of the trial-level, list-wide PC estimate and, consequently, an intermediate level of proactive control, which thus could have been high enough to prevent the need for reactive control.

Hence, our work contributed to provide consistent evidence for the specificity of the proactive control mechanism, by showing that it operated independently from the concurrent activation of reactive control and/or contingency learning. Moreover, we have provided initial evidence that cognitive control can also operate through an item-specific, PC-induced reactive control mechanism. The latter, however, emerged only in interaction with proactive control as, when assessed by itself, it could be explained mainly by contingency. This pattern thus suggests that proactive control is mainly engaged to solve Stroop interference, with a greater impact on participants' performance, as compared to item-specific, PC-induced reactive control.

Testing of A Priori Formal Models

The pattern of results discussed above suggested that, although proactive and reactive control interacted, the former had a stronger effect than the latter, as it also operated by itself. At first glance, this claim matches one alternative of our a priori hypotheses, that is, the one predicting a dominant role of proactive over reactive control. However, we tested all of our hypotheses to verify whether our data actually correlated the most with that alternative or whether another model better fitted the pattern of results we found.

The model that was more highly correlated to our data, and thus that best explained the obtained pattern of results, was the one predicting the dominance of proactive over reactive control and an antagonistic interaction between them. This result thus confirmed the asymmetrical pattern we observed in our results. Moreover, it further supported our claim that proactive control had a stronger effect than reactive control. This is in line with the fact that, in all our analyses, reactive control did not survive by itself and thus it would have been odd if our data were better explained by the model implying an equal strength of proactive and reactive control or, even more, by the model predicting the dominance of reactive control.

What is even more noteworthy here, as it emerged less straightforwardly from our results, was the direction of the three-way interaction. The fact that the best model was the one predicting an antagonistic interaction provides a better insight into how proactive and reactive control interacted to modulate the Stroop effect. This suggests that proactive and reactive control interacted in an antagonistic manner and, when both were high, they yielded a Stroop effect reduction that was smaller than that predicted by their additive effects. This means that one of the two mechanisms produced a mitigation of the strength of the effect of the other. As the stronger control mode was the proactive one, it can be assumed that it exerted such a moderator effect on reactive control, probably because this latter mechanism was not useful or effective. In other words, the possibility of relying on the trial-level, list-wide PC estimates to exert proactive control, the stronger and more effective mode, made the implementation of item-specific, PC-induced reactive control unnecessary. Figure 6A clearly displays this, as can be seen from the pattern of observed results, when proactive control was high (blue line, HighProactive/HighReactive and HighProactive/LowReactive conditions), the Stroop effect did not decrease (and was even numerically larger) when both control strategies were available (HighProactive/HighReactive) as compared to when just proactive control strategies were implemented (HighProactive/LowReactive). By contrast, when proactive control was low (red line, LowProactive/HighReactive and LowProactive/LowReactive conditions), the Stroop effect was lower when reactive control was implemented (LowProactive/HighReactive), even if still higher than that observed under proactive control, as compared to when no control mode was available (LowProactive/LowReactive).

Importantly, the observed pattern of the interaction between proactive and reactive control in modulating the Stroop effects is well in line with the DMC model proposal (De Pisapia & Braver, 2006), which assumes that reactive control preferentially operates when proactive control is not possible or advantageous. However, to the

best of our knowledge, this interplay between proactive and reactive control has never been tested directly, that is, by actually manipulating both the corresponding experimental variables, like list-wide PC and item-specific PC, respectively. Indeed, DMC model-inspired studies investigating or interpreting the behavioral and neurophysiological correlates of proactive and reactive control usually only varied list-wide PC (or list-wide PC-like) experimental variables and then assumed that reactive control operated for lower levels of proactive control. Our results thus represent a substantial contribution to the field by showing, for the first time, the actual fine-grained pattern of the proactive–reactive control interplay in modulating Stroop performance. More importantly, our results also raise severe concerns about the assumptions made by the existing studies mentioned above, namely, that when there is no or low proactive control, then reactive control must necessarily be active. Indeed, our results clearly showed that, when trial-level, list-wide PC was high (and thus proactive control was low), the Stroop effect decreased only when trial-level, item-specific PC was low (and thus reactive control was high), whereas it was very large when trial-level, item-specific PC was also high (and thus reactive control was low). In other words, high levels of list-wide PC would be a necessary but not sufficient condition for the activation of reactive control, as it might be actually implemented only when there are low levels of item-specific PC. It is important here to note that this argument is valid only for the "faster" or "associative" form of reactive control, that is, the one signaled by item-specific PC (or other forms of stimulus-related information) that, in turn, activates a sort of stimulus-attention association to apply attentional biases to stimulus-related processing in a reactive way (Bugg, 2017; Bugg & Crump, 2012; Tafuro et al., 2020).

Overall, the model testing confirmed in a more straightforward way what we obtained from our analyses. Therefore, it did not add more information to the result parameters we present in Table 3. However, we believe that it could add two main strengths to the present work: (a) The graphical representation displaying the model better showing the results provides a clearer and more easily interpretable overview of the pattern of results; (b) it encourages comparability with other works testing both similar and distinct hypotheses.

Internal Reliability

The internal reliability of our effects of interest was overall quite high for all of our analyses. In the results from the aggregated sample, the internal reliability of the effects was always higher than .73. The Stroop effect showed the highest internal reliability with an r_{SB} value of .92, and this value was even higher than those obtained in our previous study (Viviani et al., 2024), probably due to the higher number of trials used in the present study. Indeed, we previously found that the peripheral task had an internal reliability of .80, whereas here it reached an r_{SB} value of .92, and the perifoveal task had an internal reliability of .80, whereas here it reached an r_{SB} value of .94. Moreover, all of these values had low variability across randomizations.

The internal reliability of the proactive control effect was slightly lower, but still remained quite high (r_{SB} value of .73 in the aggregated sample) and was similar not only to that of the reactive control effect (r_{SB} value of .75 in the aggregated sample), but also to that of the three-way interaction (r_{SB} value of .78). This latter result is surprising, since the reliability of differential effects is generally lower.

Overall, these results confirmed our previous findings (Viviani et al., 2024), as they are still at odds with the reliability paradox (Hedge et al., 2018) and related proposals (Rouder & Haaf, 2019) according to which, if an experimental effect is large and universal, its internal reliability can hardly be large enough. As such, this further supports our choice of using multilevel modeling which, by allowing explaining intrasubject/intertrial variance, has been shown to effectively provide more precise estimates of the effects of interest.

Constraints on Generalizability

While we generally presume that the Stroop effect (and maybe its PC-dependent modulations) is a universal feature of cognition, our experiments have some constraints. Our data samples indeed largely consist of students at the University of Padova. Although we expect that our main findings will generalize across different populations of the same age range, more research should be performed to confirm this. However, our results might not be generalizable to all populations, especially not to those with attentional (or cognitive) deficits. Additionally, the age of the participants can influence the impact of cognitive control abilities on Stroop performance (e.g., Braver et al., 2001; Tafuro et al., 2019). Moreover, the results might apply only to conflict tasks that fulfill all the methodological requirements to yield a complete Stroop effect (see Viviani et al., 2023). Finally, given that our task is not tied to any culture-specific behavior (such as reading direction), we do not think that the geographic origin of our participants would influence the present results.

Conclusions

In two behavioral experiments, we aimed to investigate the DMC model and its application to the Stroop task by developing a novel methodological approach and combining it with the state-of-the-art trial-level multilevel modeling that ensures accurate and reliable estimates of the Stroop effect. This approach allowed us not only to manipulate list-wide PC and item-specific PC simultaneously, but also in a fine-grained way at the trial level, which is crucial for understanding the coexistence and interaction of proactive and reactive control, while also controlling for the confounding effects of low-level processes, including contingency.

Our results provided consistent evidence for the existence of list-wide, PC-dependent proactive control mechanisms modulating Stroop performance regardless of confounders and item-specific PC-dependent reactive control levels. Moreover, albeit we did not find evidence for the existence of specific item-specific PC-dependent reactive control effects, item-specific, PC-induced reactive control still interacted with proactive control in modulating Stroop performance, with the characteristic pattern assumed by the dual-mechanisms of control model.

Although further research is needed to validate these findings and understand the nature of the three-way interaction between congruency and proactive and reactive control, thanks to our novel approach, our study provided new insights into the DMC model and the cognitive mechanisms underpinning the modulation of the Stroop effect. These insights specifically concern three points, as detailed in what follows. First, our results reveal that using trial-level estimates of the PC provides a better account of adaptive control employment, thus encouraging future studies to do the same to reach a more realistic understanding of adaptive control modulations. Second, thanks to the simultaneous and fine-grained manipulation of list-wide PC and item-specific PC, we provided evidence for the interplay between proactive and reactive

control in modulating the Stroop effect, which sets the bases for further studies that could deepen the understanding of the two control modes postulated by the DMC model. Third, thanks to our analysis-level control approach used to control for (and estimate) the effect of contingency, we unveiled the interplay between item-specific, PC-induced reactive control and contingency, shedding light on the related still unresolved diatribe between pure associative learning and adaptive control accounts of the Stroop performance, promoting a more moderate view. Overall, our results contribute to the ongoing research on cognitive control mechanisms and their implications for understanding human cognition.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28. https://doi.org/10.21500/20112084.807

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Blais, C., & Bunge, S. (2010). Behavioral and neural evidence for item-specific performance monitoring. *Journal of Cognitive Neuroscience*, *22*(12), 2758–2767. https://doi.org/10.1162/jocn.2009.21365

Blais, C., Robidoux, S., Risko, E. F., & Besner, D. (2007). Item-specific adaptation and the conflict-monitoring hypothesis: A computational model. *Psychological Review*, *114*(4), 1076–1086. https://doi.org/10.1037/0033-295X.114.4.1076

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Braem, S., Bugg, J. M., Schmidt, J. R., Crump, M. J., 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

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. https://doi.org/10.1016/j.tics.2011.12.010

Braver, T. S., Barch, D. M., Keys, B. A., Carter, C. S., Cohen, J. D., Kaye, J. A., Janowsky, J. S., Taylor, S. F., Yesavage, J. A., Mumenthaler, M. S., Jagust, W. J., & Reed, B. R. (2001). Context processing in older adults: Evidence for a theory relating cognitive control to neurobiology in healthy aging. *Journal of Experimental Psychology: General*, *130*(4), 746–763. https://doi.org/10.1037/0096-3445.130.4.746

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In A. R. A. Conway, C. Jarrold, & M. J. Kane (Eds.), *Variation in working memory* (pp. 76–106). Oxford University Press.

Braver, T. S., Kizhner, A., Tang, R., Freund, M. C., & Etzel, J. A. (2021). The dual mechanisms of cognitive control project. *Journal of Cognitive Neuroscience*, *33*(9), 1990–2015. https://doi.org/10.1162/jocn_a_01768

Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences*, *106*(18), 7351–7356. https://doi.org/10.1073/pnas.0808187106

Bugg, J. M. (2012). Dissociating levels of cognitive control: The case of Stroop interference. *Current Directions in Psychological Science*, *21*(5), 302–309. https://doi.org/10.1177/0963721412453586

Bugg, J. M. (2014). Conflict-triggered top-down control: Default mode, last resort, or no such thing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 567–587. https://doi.org/10.1037/a0035032

Bugg, J. M. (2017). Context, conflict, and control. In T. Egner (Ed.), *The Wiley handbook of cognitive control* (pp. 79–96). Wiley Blackwell. https://doi.org/10.1002/9781118920497.ch5

Bugg, J. M., & Chanani, S. (2011). list-wide control is not entirely elusive: Evidence from picture–word Stroop. *Psychonomic Bulletin & Review*, *18*(5), 930–936. https://doi.org/10.3758/s13423-011-0112-y

Bugg, J. M., & Crump, M. J. (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*, Article 367. https://doi.org/10.3389/fpsyg.2012.00367

Bugg, J. M., & Hutchison, K. A. (2013). Converging evidence for control of color–word Stroop interference at the item level. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 433–449. https://doi.org/10.1037/a0029145

Bugg, J. M., Jacoby, L. L., & Chanani, S. (2011). Why it is too early to lose control in accounts of item-specific proportion congruency effects. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(3), 844–859. https://doi.org/10.1037/a0019957

Bugg, J. M., Jacoby, L. L., & Toth, J. P. (2008). Multiple levels of control in the Stroop task. *Memory & Cognition*, *36*(8), 1484–1494. https://doi.org/10.3758/MC.36.8.1484

Chiew, K. S., & Braver, T. S. (2017). Context processing and cognitive control: From gating models to dual mechanisms. In T. Egner (Ed.), *The Wiley handbook of cognitive control* (pp. 143–166). Wiley Blackwell. https://doi.org/10.1002/9781118920497.ch9

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Cohen, J. D. (2017). Cognitive control: Core constructs and current considerations. In T. Egner (Ed.), *The Wiley handbook of cognitive control* (pp. 3–28). Wiley Blackwell. https://doi.org/10.1002/9781118920497.ch1

De Pisapia, N., & Braver, T. S. (2006). A model of dual control mechanisms through anterior cingulate and prefrontal cortex interactions. *Neurocomputing*, *69*(10-12), 1322–1326. https://doi.org/10.1016/j.neucom.2005.12.100

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Gonthier, C., Braver, T. S., & Bugg, J. M. (2016). Dissociating proactive and reactive control in the Stroop task. *Memory & Cognition*, *44*(5), 778–788. https://doi.org/10.3758/s13421-016-0591-1

Gratton, G., Cooper, P., Fabiani, M., Carter, C. S., & Karayanidis, F. (2018). Dynamics of cognitive control: Theoretical bases, paradigms, and a view for the future. *Psychophysiology*, *55*(3), Article e13016. https://doi.org/10.1111/psyp.13016

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

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

Jacoby, L. L., Lindsay, D. S., & Hessels, S. (2003). item-specific control of automatic processes: Stroop process dissociations. *Psychonomic Bulletin & Review*, *10*(3), 638–644. https://doi.org/10.3758/BF03196526

Jiang, J., Heller, K., & Egner, T. (2014). Bayesian Modeling of flexible cognitive control. *Neuroscience & Biobehavioral Reviews*, *46*, 30–43. https://doi.org/10.1016/j.neubiorev.2014.06.001

Kornblum, S. (1992). Dimensional overlap and dimensional relevance in stimulus–response and stimulus–stimulus compatibility. In S. Kornblum (Ed.), *Tutorials in motor behavior* (Vol. 2, pp. 743–777). North-Holland.

Lindsay, D. S., & Jacoby, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 219–234. https://doi.org/10.1037/0096-1523.20.2.219

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

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 149–157. https://doi.org/10.1037/0278-7393.16.1.149

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203. https://doi.org/10.1037/0033-2909.109.2.163

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, *5*, Article 39. https://doi.org/10.3389/fnhum.2011.00039

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*(1), 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Nee, D. E., Wager, T. D., & Jonides, J. (2007). Interference resolution: Insights from a meta-analysis of neuroimaging tasks. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(1), 1–17. https://doi.org/10.3758/CABN.7.1.1

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Parr, T., Holmes, E., Friston, K. J., & Pezzulo, G. (2023). Cognitive effort and active inference. *Neuropsychologia*, *184*, Article 108562. https://doi.org/10.1016/j.neuropsychologia.2023.108562

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, *26*(2), 452–467. https://doi.org/10.3758/s13423-018-1558-y

Schmidt, J. R. (2013). Questioning conflict adaptation: Proportion congruent and Gratton effects reconsidered. *Psychonomic Bulletin & Review*, *20*(4), 615–630. https://doi.org/10.3758/s13423-012-0373-0

Schmidt, J. R. (2019). Evidence against conflict monitoring and adaptation: An updated review. *Psychonomic Bulletin & Review*, *26*(3), 753–771. https://doi.org/10.3758/s13423-018-1520-z

Schmidt, J. R., & Besner, D. (2008). The Stroop effect: Why proportion congruent has nothing to do with congruency and everything to do with contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(3), 514–523. https://doi.org/10.1037/0278-7393.34.3.514

Schmidt, J. R., Crump, M. J. C., Cheesman, J., & Besner, D. (2007). contingency learning without awareness: Evidence for implicit control. *Consciousness and Cognition*, *16*(2), 421–435. https://doi.org/10.1016/j.concog.2006.06.010

Spinelli, G., & Lupker, S. J. (2020). Item-specific control of attention in the Stroop task: Contingency learning is not the whole story in the item-specific proportion-congruent effect. *Memory & Cognition*, *48*(3), 426–435. https://doi.org/10.3758/s13421-019-00980-y

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

Stoet, G. (2010). Psytoolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). Psytoolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Tafuro, A., Ambrosini, E., Puccioni, O., & Vallesi, A. (2019). Brain oscillations in cognitive control: A cross-sectional study with a spatial Stroop task. *Neuropsychologia*, *133*, Article 107190. https://doi.org/10.1016/j.neuropsychologia.2019.107190

Tafuro, A., Vallesi, A., & Ambrosini, E. (2020). Cognitive brakes in interference resolution: A mouse-tracking and EEG co-registration study. *Cortex*, *133*, 188–200. https://doi.org/10.1016/j.cortex.2020.09.024

Thomas, D. R., & Zumbo, B. D. (2012). Difference scores from the point of view of reliability and repeated-measures ANOVA: In defense of difference scores for data analysis. *Educational and Psychological Measurement*, *72*(1), 37–43. https://doi.org/10.1177/0013164411409929

Torres-Quesada, M., Funes, M. J., & Lupiáñez, J. (2013). Dissociating proportion congruent and conflict adaptation effects in a Simon–Stroop procedure. *Acta Psychologica*, *142*(2), 203–210. https://doi.org/10.1016/j.actpsy.2012.11.015

van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, *38*(4), 584–589. https://doi.org/10.3758/BF03193889

Visalli, A., Ambrosini, E., Viviani, G., Sambataro, F., Tenconi, E., & Vallesi, A. (2023). On the relationship between emotions and cognitive control: Evidence from an observational study on emotional priming Stroop task. *PLoS ONE*, *18*(11), Article e0294957. https://doi.org/10.1371/journal.pone.0294957

Viviani, G., Visalli, A., Finos, L., Vallesi, A., & Ambrosini, E. (2024). A comparison between different variants of the spatial Stroop task: The influence of analytic flexibility on Stroop effect estimates and reliability. *Behavior Research Methods*, *56*(2), 934–951. https://doi.org/10.3758/s13428-023-02091-8

Viviani, G., Visalli, A., Montefinese, M., Vallesi, A., & Ambrosini, E. (2023). The Stroop legacy: A cautionary tale on methodological issues and a proposed spatial solution. *Behavior Research Methods*. Advance online publication. https://doi.org/10.3758/s13428-023-02215-0

Received October 5, 2023
Revision received February 16, 2024
Accepted March 7, 2024 ■