

This is a pre print version of the following article:

Tango of Control: The Interplay Between Proactive and Reactive Control / Viviani, G.; Visalli, A.; Montefinese, M.; Vallesi, A.; Ambrosini, E.. - In: JOURNAL OF EXPERIMENTAL PSYCHOLOGY. GENERAL. - ISSN 0096-3445. - 153:6(2024), pp. 1644-1670. [10.1037/xge0001585]

Terms of use:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

08/05/2026 14:57

(Article begins on next page)

Tango of Control: The Interplay between Proactive and Reactive Control

Giada Viviani^{1,2}, Antonino Visalli³, Maria Montefinese⁴, Antonino Vallesi^{1,2}, Ettore Ambrosini^{1,2,5*}

¹ Department of Neuroscience, University of Padova, Padova, Italy

² Padova Neuroscience Center, University of Padova, Padova, Italy

³ IRCCS San Camillo Hospital, Venice, Italy

⁴ Department of Developmental and Social Psychology, University of Padova, Padova, Italy

⁵ Department of General Psychology, University of Padova, Padova, Italy

*Corresponding author:

Ettore Ambrosini,

Department of Neuroscience, University of Padova, Padova, Italy

email: ettore.ambrosini@unipd.it

This unpublished manuscript is submitted for peer review

Abstract

Cognitive control has been theorized operating through two distinct mechanisms, proactive and reactive control, as posited by the Dual Mechanism of Control Model (DMC). Despite the DMC potential to explain cognitive control variability, the supporting evidence remains inconclusive. Prior studies frequently employed the Stroop task to assess the DMC, manipulating the proportion of congruency (PC) at the list-wide (LWPC) and/or item-specific (ISPC) 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 the Peripheral and the Perifoveal spatial Stroop tasks, 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 LWPC and ISPC at the trial level, allowing for a fine-grained analysis.

Our results provide compelling evidence for the existence of an LWPC-dependent proactive control mechanism, influencing Stroop performance independently of reactive control and confounding factors. Additionally, an ISPC-dependent reactive control effect was found to interact with proactive control, influencing Stroop performance, but not directly.

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.

1. Introduction

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 Mechanisms 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, 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.

1.1. 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 proportion congruency (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 mostly incongruent blocks (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 mostly congruent blocks (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 Dual Mechanisms of Control model (DMC, Braver, 2012; Braver et al., 2007). As introduced above, the DMC 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.

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 (LWPC) and item-specific (ISPC) levels (Bugg, 2012; Bugg & Crump, 2012)¹. Indeed, although these PC manipulations¹ 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 LWPC 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 LWPC, in which the proportion of congruent trials is higher (e.g., 75% congruent and 25% incongruent), are compared to

¹ A third type of PC manipulation exists, the context-specific PC. Besides being akin to the ISPC, 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.

blocks with low LWPC, wherein the ratio is reversed (e.g., 25% congruent and 75% incongruent). Typically, this manipulation yields the so-called LWPC effect, characterized by smaller Stroop effects in low-LWPC blocks as compared to high-LWPC blocks (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 LWPC 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 LWPC 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-LWPC blocks, yielding not only shorter RTs on incongruent trials, but also a congruency cost, namely a slowing on congruent trials as compared to congruent trials in high-LWPC 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 ISPC 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, the 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: high-ISPC (e.g., 75% probability for the word RED to appear in red ink) and low-ISPC (e.g., 25% probability for the word BLUE to occur in blue ink) items. Therefore, low-ISPC 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 ISPC effect, that is, smaller Stroop effects for low- than high-ISPC 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 (LWPC 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 ISPC 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 et al., 2007; Schmidt & Besner, 2008). 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 et al., 2007; Schmidt & Besner, 2008).

The contingency hypothesis challenges especially the ISPC effect, pointing out that such effect is only incidental, since ISPC 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-ISPC 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 and colleagues (2011), the task-relevant dimension signaled the ISPC rather than the task-irrelevant one and this, by equating contingency across conditions, allowed deconfounding ISPC and contingency and finding evidence for a control modulation. When, instead, the task-irrelevant dimension functioned as the ISPC signal, contingency was confounded with ISPC and accounted for its effect, as predicted by the contingency hypothesis (Schmidt, 2019). Moreover, Bugg and Hutchison (2013, Experiment 3), restoring the traditional ISPC design (the task-irrelevant dimension signaling the ISPC), found an ISPC-dependent control modulation when 4-item sets were used, while contingency was dominant when 2-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 ISPC 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 ISPC 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 ISPC in a partially independent way. When using a color-word Stroop task, this can be done by creating two non-overlapping 2-item-sized sets, so to have: i) the first set with mostly-congruent (MC) words, but mostly-incongruent (MI) colors (viceversa for the second set); ii) MC incongruent words presented only in the other MC colors; iii) MI incongruent words presented only in the other MI colors; iv) for each set, one high-contingency and one low-contingency incongruent word. To measure conflict adaptation effects, low-contingency MC incongruent items are compared to low-

contingency MI incongruent items, whereas to measure contingency-related effects, low-contingency MI incongruent items are contrasted to high-contingency MI incongruent items. Of note, for congruent items, this does not dissociate between the two accounts (e.g., Spinelli & Lupker, 2020). Recently, Braem and colleagues (2019) summarized some guidelines on how to design cognitive control tasks avoiding confounds and, for what concerns ISPC manipulations, they suggest creating two sets of overlapping “inducer” items (one for MI and one for MC items) to trigger reactive control and a third set of “diagnostic” items with a PC of 50%, to which the ISPC-dependent level of cognitive control is assumed to be transferred, to measure the ISPC effect without item-frequency differences. Using the picture-word Stroop task, diagnostic items can be created by choosing a novel set of exemplars of the inducer items which can represent the same inducer item but in a different form (e.g., if among the inducer items there is a picture of a dog, select a different dog picture as a diagnostic item) or by selecting an alternative item belonging to the same category as the inducer item (e.g., if among the inducer items there is a picture of a lion, select a tiger picture as a diagnostic item). Alternatively, they propose to use diagnostic items involving new task-irrelevant features, such as diagnostic trials with the same task-relevant features as the MC and MI items paired equally often with incongruent non-inducer task-irrelevant features (e.g., different distractor words in the color-word Stroop task). Faster responses for MI inducer task-relevant features paired with non-inducer incongruent task-irrelevant features than for MC inducer task-relevant features paired with non-inducer incongruent task-irrelevant features are assumed to reflect a cognitive control-driven ISPC effect without contingency confounds. It is worth mentioning that although these solutions have been proposed in a consensus paper, the same authors admit that they come with cautionary notes. Indeed, they recommend the latter approach 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 LWPC manipulation. Some authors indeed pointed out that LWPC 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 LWPC manipulation (Blais et al., 2007; Blais & Bunge, 2010; Bugg et al., 2008). Indeed, when low-PC blocks are composed of low-PC items and high-PC

blocks are composed of high-PC items, LWPC is confounded with ISPC. To disentangle the two mechanisms, Bugg and Chanani (2011) randomly intermixed, in both high- and low-PC blocks, an additional set of items with an ISPC of 50% (unbiased or diagnostic items) to verify whether an LWPC effect could be observed for such items, which did not provide any item-specific or contingency information. This was the case, suggesting that the LWPC 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 and colleagues (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 LWPC and ISPC 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 ISPC manipulation specifically, the creation of multiple sets of stimuli consisting of a multitude of items is not always feasible as, 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, contingency-control manipulations in picture-word Stroop tasks might also be flawed, because when using different diagnostic pictures of the same exemplars of inducer items, response contingencies might still be predictable, while when using diagnostic pictures of the different exemplars but belonging to the same category as inducer items implies the assumption of within-category transfer, which might not always occur. 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. As an alternative to this approach, there is the possibility of controlling 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 and colleagues (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 LWPC and ISPC effects in the same participants. To this end, separate blocks were used: two LWPC blocks (one mostly congruent and one mostly incongruent) along with a set of unbiased items to avoid ISPC-related influence, and one ISPC block including an equal number of mostly congruent and mostly incongruent items, with a LWPC of 50% to exclude LWPC effects. They found not only that LWPC and ISPC 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 and colleagues (2016) provided initial evidence for the separability of LWPC and ISPC 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 LWPC 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 ISPC, leading to unequal frequencies of unique trial types and irrelevant stimulus characteristics. A second limitation of the study by Gonthier and colleagues (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 LWPC and ISPC 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.

1.2. 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 LWPC and ISPC 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 of them 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 LWPC and ISPC 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, LWPC and ISPC, which can tell us more about the impact of variable amounts of proactive and reactive control activated by different levels of LWPC and ISPC. 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 LWPC and ISPC, with the aim to estimate its effect and control for it at the statistical level so to measure the LWPC and ISPC effects regardless of contingency. As a consequence, in our tasks, trials will have at the same time a different level of LWPC, ISPC, 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 partialling out the effect of contingency and of other lower-level confounding factors (Viviani et al., 2023b). 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 (GLM), 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., 2023b for a more exhaustive description of the advantages of multilevel modeling over GLM).

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 LWPC and ISPC 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 LWPC and ISPC 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 LWPC and ISPC 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 LWPC value at, say, the 20th trial in a block deviates even dramatically from the expected block-level LWPC, being it, for example, as large as 40% and as small as 0% instead of 20%. This is especially important after an unsignalled block transition, especially between blocks with extreme opposite block-level LWPCs (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 LWPC estimates (from 20% to 80% in this example) and, consequently, their proactive control level. Similarly, commonly used block-level ISPC values neglect the fact that participants first need to experience a sufficient number of trials for each item to estimate its ISPC value, thus unrealistically assuming that the items at the beginning of each block have already been associated to an ISPC 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 LWPC and ISPC 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 LWPC and ISPC, 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.

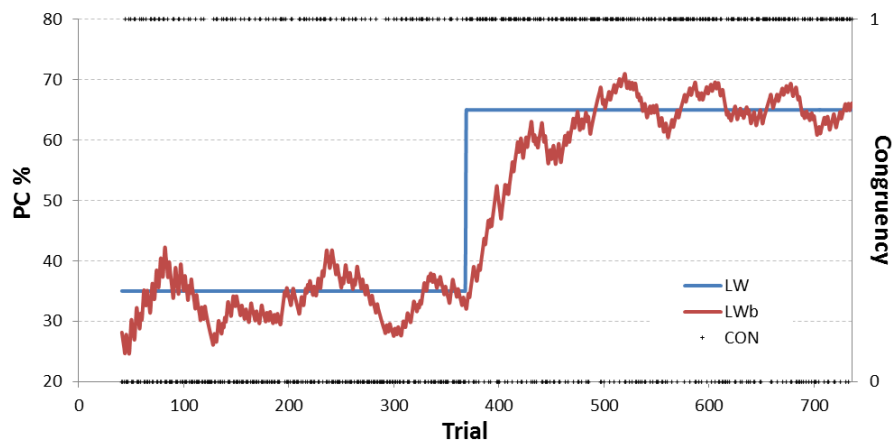


Figure 1

The plot shows the block-level LWPC (LW, blue line) and its trial-level estimates (LWb, red 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 plus signs.

Finally, we addressed the important but frequently overlooked aspect of measure reliability. As highlighted by Gonthier and colleagues (2016), LWPC and ISPC effect indices have unknown psychometric properties, in addition to being effects 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 GLM approaches (see Viviani et al., 2023b).

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., 2023b).

1.3. 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 LWPC and ISPC 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 LWPC and ISPC are high (IPro-IRea 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 LWPC or ISPC are low, a high level of proactive and reactive control, respectively, should be implemented (respectively, the

hPro-IRea and hRea-IPro conditions), and thus the Stroop effect should be reduced. Finally, when both LWPC and ISPC are low (hPro-hRea 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 hPro-hRea 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 (ADD 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 (SYN 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 (ANT models; Figure 2C).

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, hPro-IRea and IPro-hRea), while it should not affect the size of the Stroop effect in the conditions wherein neither or both forms of control are implemented (respectively, IPro-IRea and hPro-hRea). If we assume that proactive and reactive controls have the same strength, the Stroop effect should be the same size in the IPro-IRea and hPro-hRea 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 hPro-IRea condition compared to the IPro-hRea condition (Figure 2, right plots), while if the effect of reactive control is stronger, the Stroop effect should be smaller in the IPro-hRea condition compared to the hPro-IRea condition (Figure 2, middle plots).

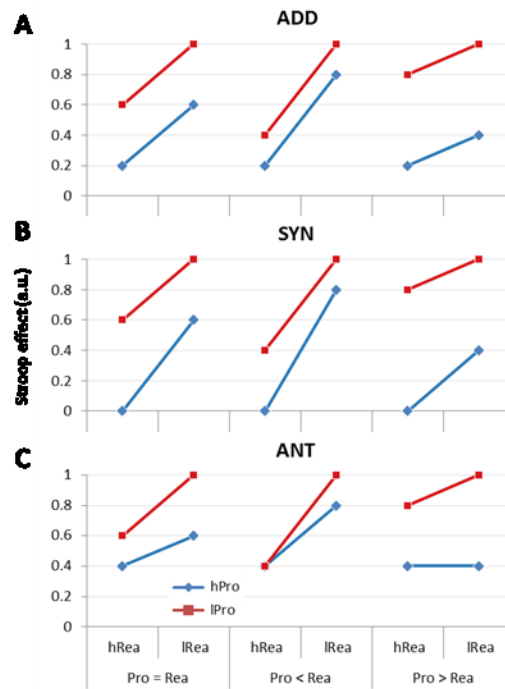


Figure 2

Predicted patterns of Stroop effect modulations by low and high levels (l and h, respectively) of LWPC-related Proactive control (Pro) and ISPC-related Reactive control (Rea) according to our alternative hypotheses. ADD, additive effects of Pro and Rea; SYN, synergistic Pro by Rea interaction; ANT, antagonistic Pro by Rea interaction (see Hypotheses section for more details).

2. Experiment 1 - Peripheral

2.1. Methods

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 Open Science Framework (OSF) platform at osf.io/qmu7g. No part of the study, including the analyses, was pre-registered.

2.1.1. 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 x 600 pixels, on a gray background (RGB: 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 2000 ms. 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., 2023b), 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 x 36 pixels) presented at the center of the screen, along with four white squares (73 x 73 pixels) at the four corners of an imaginary square of 600 x 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. 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 towards the correct response.

List-wide (LWPC) and item-specific (ISPC) 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 LWPC values, one with 35% of congruent trials (LW35) and one with 65% of congruent trials (LW65). Then, distinct ISPC levels were used, nested within each LWPC block, so as to have 4 different ISPC values within each block. In the LW35 block, the ISPC ranged from 20% to 50% in steps of 10%, while in the LW65 block, the ISPC ranged from 50% to 80% in steps of 10%. Crucially, by using the same ISPC level (50%) in both LWPC blocks, we were able to assess the pure effect of LWPC (and thus proactive control) on Stroop effects, independently

of ISPC (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 LWPC and ISPC, 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 sub-block, 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 ISPC level, two different contingency values differing by 10% were used for the two possible incongruent trials (see Figure 3). In addition to the 640 experimental trials, before each LWPC block, we added sub-blocks of 40 trials to favor the familiarization of participants to the current block LWPC level. Moreover, the 640 experimental trials were divided into 8 sub-blocks with self-paced breaks in between, and at the beginning of each sub-block we added 2 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 LWPC, ISPC and contingency for each trial-list version using the Hierarchical Gaussian Filter (HGF, Mathys et al., 2011). The HGF is a filter that uses variational Bayes under a mean-field approximation to update the probability of an event on each trial. Specifically, trial-level estimates were computed reflecting trial-by-trial probabilities updated based on: i) the stimulus congruency, for LWPC; ii), the stimulus congruency conditional to a specific position, for ISPC; iii) the target direction (and thus the response) conditional to a specific position, for contingency. The HGF 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 LWPC and ISPC 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.

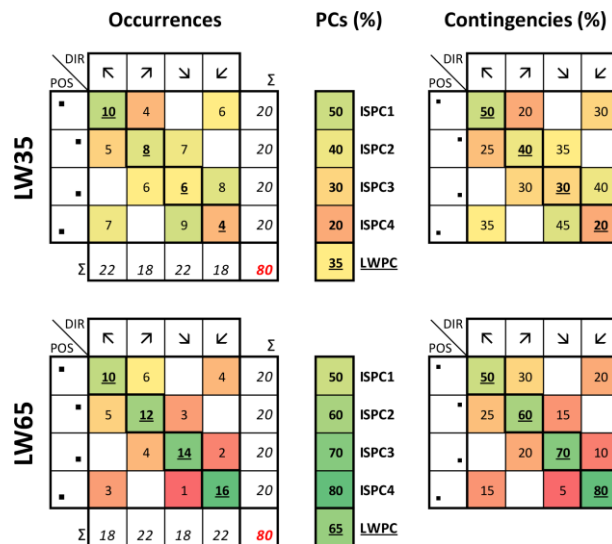


Figure 3

Block-wise structure of the task. Separately for one sub-block of each LWPC blocks (LW35, top row; LW65, bottom row), the image shows on the left the number of trials (occurrences) with a specific target direction (DIR) and position (POS). For example, in the LW35 sub-block, we had nine trials with the arrow appearing in the lower-left corner, but pointing towards the lower-right corner. The trials in the diagonal are the congruent ones (underscored). For each sub-block, the corresponding contingencies are also shown on the right, while the middle column shows the percentage of congruent trials (LWPC) and of congruent trials specific for each location (ISPC). The color scale indicates the relative probability of each trial type/contingency, as well as the relative level of the LWPCs/ISPCs.

2.1.2. Data analysis

Various analyses were conducted to estimate the effect of LWPC and ISPC 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, MA).

The analyses were performed on inverse-transformed RTs (iRTs, computed as $-1000/RTs$). This transformation was employed to eliminate the heavy right skewness of the RT distribution, for which the logarithmic transformation was not enough. From the 62720 experimental trials, we excluded trials with

incorrect responses ($n = 2115$, 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 ($n = 2103$, 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 Supplementary Materials at osf.io/aeh2d, Figure S1.2)

2.1.2.1. Assessing the magnitude of LWPC and ISPC 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 (LMM, 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 LWPC and ISPC, as well as their interaction, in modulating the Stroop effect when they both varied. Moreover, this approach allowed us to do so while partialling 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 Introduction, 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 (GLM) 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 intra-subject/inter-trial sources of variance contributing to measurement error, it also provides better reliability of these estimates (Viviani et al., 2023b).

We tested a LMM 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, indicated by the suffix “b” to the predictor names (because they were estimated using the ideal Bayesian observer). 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 i) a continuous predictor reflecting the iRT of the preceding trial (iRT_{pre}), 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; ii) the continuous predictor for the effect of trial number ($trial_{TOT}$) to account for potential time-on-task effects, such as the effects of learning/adaptation or fatigue; iii) the horizontal and vertical position of the stimulus on the screen (respectively, hS and vS), to account for potential (e.g., perceptual, attentional) differences due to the location where the stimulus appeared (left vs right, above vs below, respectively); iv) the horizontal and vertical coding of the response (respectively, hR and vR), 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, v) contingency, which is the conditional probability of the response given the stimulus, $P(R|S)$ (PRS_b), and vi) the probability of the response, $P(R)$ (PR_b). The experimental effects of interest were modeled by including the predictors for the LWPC and ISCP manipulations (LW_b and IS_b , respectively) and trial congruency (CON), 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: $RT \sim iRT_{pre} + Trial_{TOT} + hS + vS + hR + vR + PRS_b + PR_b + LW_b * IS_b * CON + (LW_b * IS_b * CON | SS)$.

The continuous predictors iRT_{pre} , $Trial_{TOT}$, PRS_b , and PR_b 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 LW_b and IS_b , 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 re-fitted a trimmed version of the model obtained by excluding data points with absolute standardized residuals exceeding 3.

We report the estimated coefficient (b), standard error (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 alpha level of .05 was set as the cut-off for statistical significance. We used the participants' random slopes to compute the individual effects of interest (that is, the Stroop effect and its modulation by LWPC, ISPC, 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 2-way interactions CON_0:LWb and CON_0:ISb), to verify whether its inclusion might have interfered with the estimation of the effects of the two 2-way interactions testing for the distinct effects of proactive and reactive control (continuous full 2-way interaction model).

To assess the robustness of our results to analytical flexibility, control analyses were also performed by replicating LMM results for the continuous full model using another multilevel modeling approach, that is, a random coefficient analysis (RCA, also called random regression or two-step regression; Lorch & Myers, 1990). For the RCA analysis, we first ran linear regressions at the subject level using the same regression model as the final LMM model described above (continuous full model). The Wilkinson-notation formula for the RCA model is: $RT \sim iRTpre + TrialTOT + hS + vS + hR + vR + PRSb + PRb + LWb*ISb*CON$.

As for the LMM 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 (GLM) 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, GLM analyses are not well-suited for our experimental design because it is incomplete (ISPC is nested in LWPC). This was not an issue for testing the Stroop effect magnitude per se, as we could aggregate congruent and incongruent trials across the LWPC and ISPC levels, but it prevented us from testing the effects of LWPC and ISPC while also controlling for contingency (see Assessing the magnitude of LWPC and ISPC effects). Indeed, due to our manipulation of LWPC, ISPC, and contingency, and since contingency is inevitably confounded with ISPC in congruent trials, we did not have all the required combinations of LWPC, ISPC, and contingency levels, and the trial number for the available combinations was heavily unbalanced. These limitations notwithstanding, we decided to perform GLM analyses anyway to favor the comparison of our results on the Stroop effect magnitude with the literature.

2.1.2.2. Assessing the internal reliability of LWPC and ISPC effects

The internal consistency of the experimental effects of interest was assessed for the LMM results by computing split-half Pearson's correlations corrected with the Spearman-Brown formula (r_{SB}). We used

2000 randomizations and calculated both the median r_{SB} values and the corresponding nonparametric 95% confidence intervals ($CI_{95\%}$).

Essentially, observations were randomly split into two subsets and, on each subset, LMM 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 partialling 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.

2.1.3. Participants

For the first experiment, we recruited 98 participants (55 females and 43 males; mean age = 25.89 years, $SD = 6.42$ years). 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 LMM analysis, the approaches available to date for power analysis are not adequate and/or feasible for our complex statistical model (see Viviani et al., 2023b for a detailed discussion), especially because it involves the interaction between continuous predictors. Nonetheless, it should be noted that the RCA and LMM approaches are quite similar and provide similar results (at least regarding the Stroop effects in our experimental paradigm), and the power analysis for RCA 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 (Erdfelder et al., 1996) 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 LWPC and ISPC) in a two-tailed one-sample t-test. We conservatively assumed a

small-medium Cohen's d effect size of .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 LMMs tend to provide higher power than standard GLM approaches like the one-sample t -test we used here.

2.2. Results

2.2.1. Magnitude of LWPC and ISPC 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 Supplementary Materials at osf.io/aeh2d, Figures S1.1 and S1.3-5).

GLM-based analyses were performed using t tests. The overall Stroop effect (i.e., collapsing across LWPC and ISPC values) was significant ($t = 35.50$, $p < .0001$) 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 Supplementary Materials at osf.io/aeh2d, Table S1.1).

Regarding the LMM analysis, we first compared the full continuous model to the full discrete model using the AIC model selection and we found that the best-fit model was the full continuous model ($AIC = 8352$ vs 8436.7 of the full discrete model). Then, we compared our full continuous model with the reduced continuous one by performing the log-likelihood ratio test, which revealed that the full continuous model was justified ($\chi^2(7) = 7530.5$, $p < .0001$), 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 LMM model was .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 Supplementary Materials at osf.io/aeh2d, Figure S1.3). This analysis revealed that all the confounding predictors were significant in modulating participants' iRTs (all $ps < .0001$, see Table 1). 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 (PRb) was higher.

Table 1 – Results of the LMM analysis for Experiment 1 (continuous full model)

Effect	<i>b</i>	<i>SE</i>	<i>t</i>	<i>df</i>	<i>p</i>
Intercept	-1.9439	0.0350	-55.559	145.60	< .0001
TrialTOT	-0.0923	0.0023	-40.030	1967.43	< .0001
CON_0	0.3521	0.0123	28.620	179.86	< .0001
iRTpre	0.0551	0.0012	47.190	57392.28	< .0001
hS	-0.0104	0.0028	-3.700	12850.75	.0002
vS	-0.0791	0.0027	-29.056	19753.88	< .0001
hR	-0.0351	0.0026	-13.381	34201.64	< .0001
vR	-0.1013	0.0026	-39.379	39671.65	< .0001
LWb	-0.0086	0.0104	-0.825	97.51	.4112
ISb	-0.0135	0.0074	-1.827	435.59	.0684
PRsb	-0.0207	0.0059	-3.497	48890.40	.0005
PRb	-0.1234	0.0133	-9.256	49487.44	< .0001
CON_0:LWb	0.0651	0.0081	8.071	101.81	< .0001
CON_0:ISb	0.0162	0.0102	1.590	939.28	.1121
LWb:ISb	-0.0123	0.0109	-1.132	97.28	.2606
CON_0:LWb:ISb	0.0140	0.0120	1.167	94.08	.2461

Notes: *b*, coefficient estimates; *SE*, standard error, *df*, degrees of freedom computed with the Satterthwaite's approximation. See the main text for the spelling out of the acronyms for the effects

For what concerns our predictors of interest, we found that the Stroop effect (CON_0) 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 effect was significantly modulated by LWb ($p < .0001$, $d_r = 1.27$, $d_s = 0.80$, dominance = 92.86%), showing that as LWb increased, the Stroop effect increased, revealing the effect of proactive control. By contrast, the LMM analysis did not reveal a significant modulation of the Stroop effect by ISb ($p = .1121$, d_r

= 0.62, $d_s = 0.05$, dominance = 70.41%). Moreover, the three-way interaction between the Stroop effect, LWb and ISb was not significant ($p = .2461$, $d_r = 0.16$, $d_s = 0.12$, dominance = 57.14%), suggesting that LWb and ISb did not interact in modulating the Stroop effect. Lastly, our analysis revealed a significant effect of Contingency (PRsb, $p = .0005$), indicating that participants responded faster when PRsb was higher.

To verify whether the effect of the CON_0:ISb interaction was hindered by the three-way interaction, we also performed the continuous full 2-way model, excluding the three-way interaction and keeping the two double interactions (CON_0:LWb and CON_0:ISb). This analysis confirmed the results reported above, and the interaction between ISb and CON_0 remained non-significant ($p = .0920$, $d_r = 0.50$, $d_s = 0.06$, dominance = 72.45%), confirming that in the Peripheral task, we did not find a significant modulation of the Stroop effect by ISb. Lastly, the effect of PRsb was again significant ($p = .0004$), showing that participants responded faster when PRsb was higher (see Supplementary Materials at osf.io/ae2d, Table S1.7).

To confirm these results, RCA 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 LMM analyses were confirmed (all p s < .004, see Supplementary Materials at osf.io/ae2d, Table S1.8), with the exception of hS. RCA results regarding our predictors of interest partially replicated LMM 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 LWb ($p < .0001$, $d_r = .86$, dominance = 77.55%). Moreover, the interaction between CON and ISb 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 LMM ones. Indeed, we found that LWb and ISb interacted significantly in modulating the Stroop effect ($p = .0009$, $d_r = 0.35$, dominance = 65.31%). Lastly, the effect of PRsb was significant ($p = .0001$, $d_r = -0.40$), confirming previous results.

Since we found that the effect of PRsb was always significant while the effect of reactive control was never significant, we hypothesized that we did not find it because PRsb 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 PRsb (continuous full

No_PRS model) both using LMM and RCA (see Supplementary Materials at osf.io/ae2hd, Tables S1.9-10).

These analyses confirmed the results of the previous ones, except for the fact that, by removing PRSb, the interaction between CON and ISb became significant (both $ps < .0001$, $d_s = 0.84$ and $d_r = 0.83$, respectively).

Of note, the inclusion of PRSb in the model was justified and improved the model fit ($\chi^2(1) = 30.1$, $p < .0001$).

2.2.2. Internal reliability of LWPC and ISPC effects

We assessed the internal reliability of our effects of interest using LMM to explain intra-subject/inter-trial 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 $CI_{95\%}$ of .89-.94. The internal reliability of the proactive control effect had a median r_{SB} value of .79 and a $CI_{95\%}$ of .67-.87, while the median r_{SB} internal reliability of the reactive control effect was .58 with a $CI_{95\%}$ 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 $CI_{95\%}$ of .65-.83 (see Figure 4 and Supplementary Materials at osf.io/ae2hd, Figure S1.9).

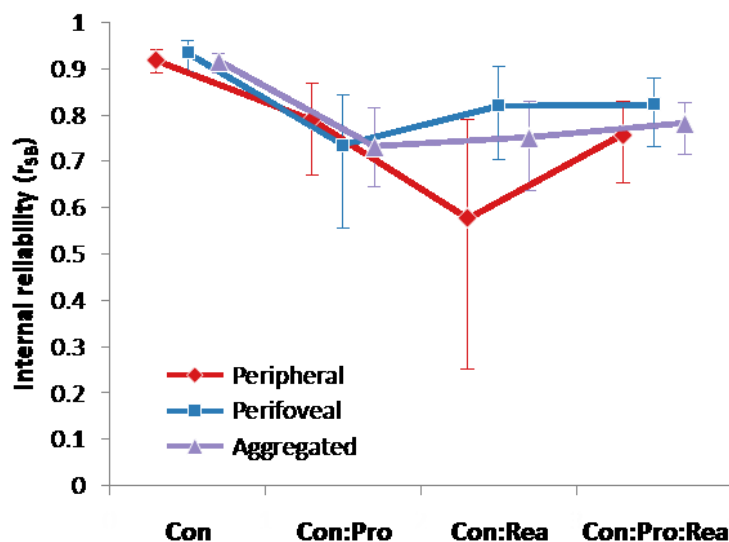


Figure 4

Internal reliability (r_{SB}) of the experimental effects of interest (Con, Stroop effect; Con:Pro, LWb modulation of Stroop effects) in the three analyses (Con:Rea, ISb modulation of Stroop effects; Con:Pro:Rea, interaction between LWb and ISb in modulating the Stroop effects). Error bars represent the non-parametric 95% confidence intervals.

2.3. Discussion

The results of our previous study (Viviani et al., 2023b) 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., 2023b), 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, among which contingency.

The analysis assessing the magnitude of such effects revealed different results based on the analytical approach employed. As explained in the methods, the standard GLM-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 RCA-based analysis but not in the LMM-based one, casting shadows on the robustness of such an effect.

A further consistent aspect revealed by our analyses is the relation between reactive control and PRSb. Indeed, they interestingly revealed that when PRSb was statistically controlled for, the ISb-dependent modulation of the Stroop effect was not significant, thus confirming the strong influence of contingency on ISPC-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 PRSb 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 4).

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., 2023b) 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., 2023b). 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.

3. Experiment 2 - Perifoveal

3.1. Methods

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 Open Science Framework (OSF) platform at osf.io/qmu7g. No part of the study, including the analyses, was pre-registered.

3.1.1. Procedure and experimental tasks

In this experiment, we used a Perifoveal spatial Stroop task (Viviani et al. 2023b), 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 x 30 pixels) enclosed in the partial outline of a black square (94 x 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.

3.1.2. 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 sub-block. From the resulting 49920 experimental trials, we also excluded trials with incorrect responses ($n = 1885$, 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 ($n = 1667$, 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 Supplementary Materials at osf.io/aeh2d, Figure S2.2)

3.1.3. Participants

For this Experiment, we recruited 78 participants (41 females and 37 males; mean age = 24.21 years, $SD = 6.43$ years). 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 LMMs tend to provide higher power than standard GLM approaches like the one-sample t test we used here.

3.2. Results

3.2.1. Magnitude of LWPC and ISPC 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 Supplementary Materials at osf.io/aeH2d, Figures S2.1 and S2.3-5).

GLM-based analyses using t tests were first used to explore the overall Stroop effect (i.e., aggregating across LWPC and ISPC values). These analyses revealed that participants responded significantly slower to Incongruent as compared to Congruent trials ($t = 33.43$, $p < .0001$) with a very large effect size ($d = 3.79$) and 100% dominance (see Supplementary Materials at osf.io/aeH2d, Table S2.1).

Then, we performed analyses using both the continuous and the discrete full models. We found that the former provided a better fit to the data ($AIC = 29422$ vs. 29511 of the full discrete model). Then, we compared the continuous full model with the continuous reduced model performing the log-likelihood ratio test, which revealed that the full continuous model was justified ($\chi^2(7) = 2971.6$, $p < .0001$), 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 LMM model was .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 Supplementary Materials at osf.io/aeH2d, Figure S2.3). All the confounding predictors of our continuous full model significantly modulated participants' iRTs (all $ps < .04$, see Table 2). 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 (PRb) was higher.

Table 2 – Results of the LMM analysis for Experiment 2 (continuous full model)

Effect	<i>b</i>	<i>SE</i>	<i>t</i>	<i>df</i>	<i>p</i>
Intercept	-2.4845	0.0448	-55.449	128.21	< .0001
TrialTOT	-0.1018	0.0034	-29.687	2686.32	< .0001
CON_0	0.3729	0.0177	21.059	138.27	< .0001
iRTpre	0.0586	0.0017	35.077	45622.01	< .0001
hS	-0.0080	0.0039	-2.072	4301.02	0.0383
vS	0.0110	0.0038	2.888	3793.44	0.0039
hR	-0.0608	0.0038	-16.143	31342.78	< .0001
vR	-0.0723	0.0037	-19.484	29671.40	< .0001
LWb	-0.0002	0.0145	-0.012	78.89	0.9907
ISb	-0.0025	0.0112	-0.223	262.95	0.8238
PRsb	-0.0516	0.0085	-6.062	40688.90	< .0001
PRb	-0.1572	0.0189	-8.300	39539.08	< .0001
CON_0:LWb	0.0668	0.0119	5.606	77.49	< .0001
CON_0:ISb	-0.0066	0.0157	-0.416	422.46	0.6774
LWb:ISb	-0.0311	0.0149	-2.085	78.49	0.0403
CON_0:LWb:ISb	0.0454	0.0204	2.226	78.29	0.0289

Notes: *b*, coefficient estimates; *SE*, standard error, *df*, degrees of freedom computed with the Satterthwaite's approximation. See the main text for the spelling out of the acronyms for the effects.

For what concerns our predictors of interest, the Stroop effect (CON_0) 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, LWPC 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 LWb increased. By contrast, the Stroop effect was not significantly modulated by ISb ($p =$

.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 LWb and ISb 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 Contingency (PRsb) was again significant ($p < .0001$), indicating that participants responded faster when PRsb was higher. As opposed to the analysis on the Peripheral task, here we do not report the continuous full 2-way model, since the three-way interaction was significant (but see Supplementary Materials at osf.io/ae2d, Table S2.2).

We then performed an RCA analysis on the continuous full model, which basically confirmed the results of the LMM analysis. Indeed, all the confounding predictors were in the same direction as the LMM results, and all but hS and vS 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 LWb ($p < .0001$, $d_r = 0.60$, dominance = 71.79%). In line with all previous results, the CON_0:ISb interaction was not significant ($p = .7178$, $d_r = -0.04$, dominance = 51.28%), showing no reactive control effect on the Stroop effect. The RCA analysis also confirmed the three-way interaction found with the LMM analysis on the Perifoveal task, showing that LWb and ISb significantly interacted in modulating the Stroop effect ($p = .0006$, $d_r = 0.41$, dominance = 65.38%). Also, the effect of PRsb was again significant as in all our analyses ($p = .0001$, $d_r = -0.47$) (see Supplementary Materials at osf.io/ae2d, Table S2.8).

Lastly, we performed the same control analysis described for the Peripheral Experiment to verify whether the non-significant effect of reactive control was due to the effect of PRsb by rerunning the continuous full model using LMM and RCA but excluding PRsb (continuous full No_PRs model). The results were confirmed but, again, the interaction between CON_0 and ISb became significant after removing PRsb (both $ps < .0001$, $d_s = 0.71$ and $d_r = 0.74$, respectively). Of note, the inclusion of PRsb in the model was justified and improved the model fit ($\chi^2(1) = 18.0$, $p < .0001$) (see Supplementary Materials at osf.io/ae2d, Tables S2.9-10).

3.2.2. Internal reliability of LWPC and ISPC 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 $CI_{95\%}$ of .89-.96. The internal reliability of proactive control

had a median r_{SB} value of .74 and a $CI_{95\%}$ of .56-.84, while the median r_{SB} internal reliability of reactive control was .82 with a $CI_{95\%}$ 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 $CI_{95\%}$ of .73-.88 (see Figure 4 and Supplementary Materials at osf.io/aeH2d, Figure S2.9).

3.3. Discussion

As discussed in our previous study (Viviani et al., 2023b) and in the Peripheral Experiment discussion (see Section 2.3), 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., 2023b), 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 ISPC-related reactive control employment, provided further

evidence for our hypothesis that, when PRSb was included in the model, there was no residual variance left to be explained by ISb. Indeed, after PRSb was removed from the model, the effect of ISPC-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 4).

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.

4. Between-Experiments comparisons

4.1. Methods

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 Open Science Framework (OSF) platform at osf.io/qmu7g. The analyses were not pre-registered.

4.1.1. Data analysis

4.1.1.1. Between-Experiment differences in LWPC and ISPC 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 Supplementary Materials at osf.io/ae2d, Figure S3.2)

First, a GLM 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 LMM 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 hS and vS (but not with hR and vR since the response effectors were the same), and with PRS 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_btww4 model, is: $RT \sim iRTpre + TrialTOT + hR + vR + PRb + Experiment*(hS + vS + PRSb + LWb*ISb*CON) + (LWb*ISb*CON|SS)$.

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 RCA 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: $RT \sim iRTpre + TrialTOT + hS + vS + hR + vR + PRSb + PRb + LWb*ISb*CON$.

After excluding outliers exceeding 3 SD , 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 datasets together, it would provide an overall result based on a larger sample.

4.1.1.2. Assessing the internal reliability of LWPC and ISPC effects

The internal consistency of the experimental effects of interest was assessed for the LMM results in the aggregate sample as described in Experiment 1.

4.2. Results

4.2.1. Between-Experiments differences in LWPC and ISPC effects

We report here only the results of the analyses performed on iRTs for the same reasons explained above.

The GLM-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 ($M_{\text{diff}} = .0832$, $t = 4.77$, $p < .0001$, $d = 0.72$).

Regarding LMM analysis, the continuous full_bt看4 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 Supplementary Materials at osf.io/aeh2d, Table S3.1). 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 PRb 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 LWb and ISb 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 PRSb was significantly different between the two Experiments ($p = .0002$), with a greater effect of PRSb 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 (CON_0), 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, LWb 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 ISb modulation of the Stroop effect ($p = .5849$, $d_r = 0.12$, $d_s = 0.02$, dominance 55.68%). Interestingly, the three-way interaction between LWb, ISb and CON_0 was significant ($p = .0094$, $d_r = 0.25$, $d_s = 0.20$, dominance 59.09%), suggesting that LWb and ISb 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 PRSb was again significant ($p < .0001$), indicating that participants responded faster when PRS was higher (see Supplementary Materials at osf.io/aeht2d, Table S3.1).

As explained in the methods section, we performed the same model but after excluding the interactions between Experiment and the predictors that resulted non-significant 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_btw model), is basically identical to the ones performed on the two Experiments separately, but here it was run on the two datasets 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_btw4 model. The conditional R^2 of the model was .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 Supplementary Materials at osf.io/aeht2d, FigureS3.3). 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 3). 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. LWb 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 ISb 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 LWb, ISb and CON_0 was again significant ($p = .0150$, $d_r = 0.24$, $d_s = 0.19$, dominance = 55.68%), indicating that LWb and ISb 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_btw4 model results.

Table 3 – Results of the LMM analysis for the btw-studies analysis (continuous full_btw model)

Effect	b	SE	t	df	p
Intercept	-2.2029	0.0278	-79.225	273.92	< .0001
TrialTOT	-0.0967	0.0020	-48.649	5466.32	< .0001
CON_0	0.3612	0.0105	34.470	311.05	< .0001
iRTpre	0.0568	0.0010	58.150	102923.55	< .0001
hS	-0.0103	0.0023	-4.394	17358.21	< .0001
vS	-0.0330	0.0023	-14.442	20631.10	< .0001
hR	-0.0463	0.0022	-21.001	72402.04	< .0001
vR	-0.0899	0.0022	-41.542	74362.37	< .0001
LWb	-0.0057	0.0087	-0.653	175.99	0.5144
ISb	-0.0083	0.0064	-1.291	656.11	0.1972
PRsb	-0.0356	0.0050	-7.152	91065.10	< .0001
PRb	-0.1335	0.0111	-11.977	92613.89	< .0001
CON_0:LWb	0.0658	0.0069	9.590	177.58	< .0001
CON_0:ISb	0.0063	0.0089	0.708	1198.54	0.4792
LWb:ISb	-0.0214	0.0090	-2.389	175.94	0.0180
CON_0:LWb:ISb	0.0279	0.0113	2.458	174.02	0.0150
Exp	-0.4712	0.0448	-10.523	182.63	< .0001
hS:Exp	-0.0098	0.0041	-2.376	17548.36	0.0175
vS:Exp	0.1047	0.0041	25.703	22145.14	< .0001
PRsb:Exp	-0.0229	0.0053	-4.357	336.10	< .0001

Notes: *b*, coefficient estimates; *SE*, standard error, *df*, degrees of freedom computed with the Satterthwaite's approximation. See the main text for the spelling out of the acronyms for the effects

RCA analysis was then performed to confirm between-Experiments LMM results and this was the case.

Indeed, all the predictors of interest were not significantly different in the two Experiments (all $ps > .14$, see Supplementary Materials at osf.io/aeH2d, Table S3.3). Therefore, this analysis confirmed the robustness of

the results obtained with RCA. Confounding predictors were all significant (all p s < .04, see Supplementary Materials at osf.io/aeH2d, Table S3.3) and in the same direction as the previous results, except for vS, 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 LWb ($p < .0001$, $d_r = 0.71$, dominance = 75%) but not by ISb ($p = .8147$, $d_r = -0.02$, dominance = 48.86%). Moreover, LWb and ISb interacted significantly in modulating the Stroop effect ($p < .0001$, $d_r = 0.37$, dominance = 65.34%), which is consistent with the RCA results from both Experiments. The effect of PRSb was also significant ($p < .0001$, $d_r = -0.43$) (see Supplementary Materials at osf.io/aeH2d, Table S3.3).

Given that between-Experiments results also showed that PRSb was significant but the CON_0 by ISb interaction was not, we further tested our hypothesis that we did not find it because PRSb explained all the variance that could have been explained by the reactive control modulation of the Stroop effect. For both the LMM and the RCA analyses, the continuous full_bt看 No_PRS model showed the same pattern of results, both for confounders and predictors of interest, with the only exception that, after removing PRSb from the model, the interaction between CON_0 and ISb became significant (p s < .0001) and with large effect sizes ($d_r = 0.72$ and 0.76 , respectively; see Supplementary Materials at osf.io/aeH2d, Tables S3.5-6). Of note, the inclusion of PRSb in the model was justified and improved the model fit ($\chi^2(2) = 138.4$, $p < .0001$).

4.2.2. 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 LMM analysis (Figure 5A) 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 5B for all correlations).

4.2.3. Internal reliability of LWPC and ISPC 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 $CI_{95\%}$ of .89-93. The internal reliability of proactive control had a median r_{SB} value of .73 and a $CI_{95\%}$ of .64-.82, while the median r_{SB} of reactive control was .75, with a $CI_{95\%}$ of .64-.83. Finally, the internal reliability of the three-way interaction was similar, with a median r_{SB} value of .78 and a $CI_{95\%}$ of .71-.83 (see Figure 4).

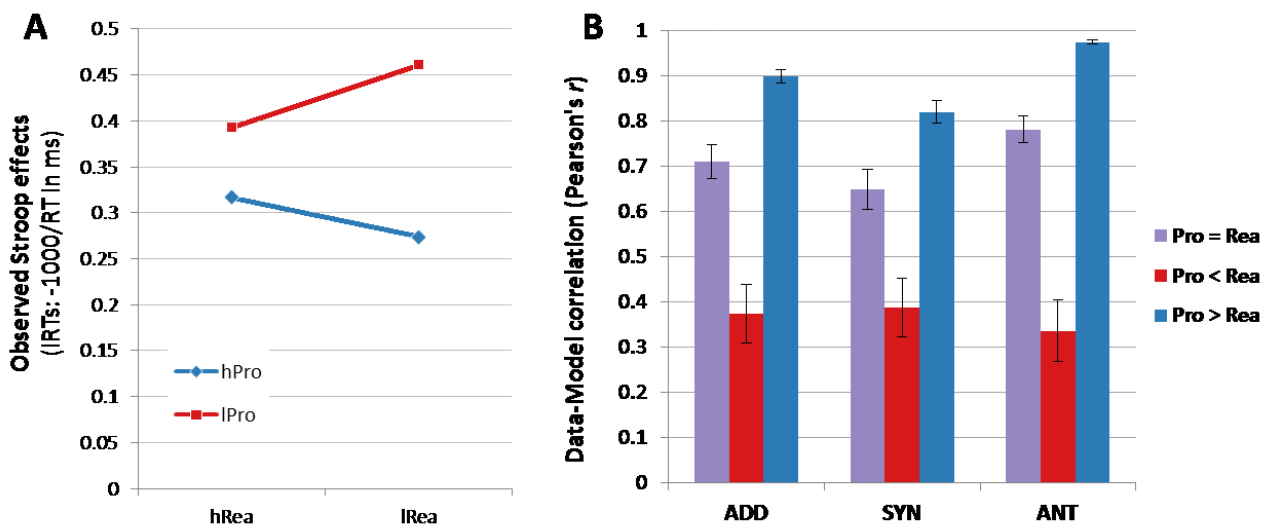


Figure 5

A) The plot shows the observed pattern of Stroop effects, as estimated by the LMM continuous full_bt看 model, as a function of the level of proactive control (hPro, blue line, corresponding to low-LWb conditions; IPro, blue line, corresponding to high-LWb conditions) and reactive control (hRea, corresponding to low-ISb conditions; IRea, corresponding to high-ISb 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. ADD, additive effects models; SYN, synergistic effects models; ANT, antagonistic effects models; Pro, Proactive control effect; Rea, Reactive control effect.

5. General discussion

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

has been frequently assessed with the Stroop task, which allows varying the Proportion of Congruency (PC) at the list-wide (LWPC) and/or at the item-specific (ISPC) 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 et al., 2007; Schmidt & Besner, 2008). 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., 2023b), with a novel methodological approach allowing to manipulate both LWPC and ISPC 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 LWPC manipulations modulated the Stroop effect, whereas ISPC-induced reactive control did not, probably due to the confounding effect of contingency. However, both control modes interacted in modulating the Stroop effect.

5.1. 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 LWPC and ISPC were manipulated simultaneously. This is indeed a more effective way to i) directly explore whether proactive and reactive control can coexist, that is, whether each mechanism can be active while also the other is activated; ii) 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. Moreover, to control for contingency-related effects, we did not balance stimulus-response combinations so as to make the contingency orthogonal to LWPC and ISPC 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 GLM-based approaches. As detailed in the Methods sections, GLM 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 partialled 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., 2023b), 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 intra-subject/inter-trial sources of variance, that 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 Section 1.1), they can hardly be properly controlled at the design level only. Indeed, it should be noted that it is practically impossible to de-confound ISPC 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 ISPC-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.

5.2. LWPC and ISPC 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, LWPC and ISPC), 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 task² 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., 2023b), 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

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.

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 RCA analyses which served us to confirm LMM results. Indeed, the RCA 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 LMM 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 proportion congruency and the Stroop effect (ISb and CON_0). The absence of such interaction was observed also in one of the control analyses performed on the Peripheral Experiment using the LMM 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 ISPC-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 contingency predictor. This was done because there is extensive literature showing that reactive control is specifically confused with contingency when ISPC is manipulated (e.g., Schmidt, 2019; Schmidt & Besner, 2008). Moreover, as explained above, it is very difficult to totally decorrelate contingency and ISPC at the design level while also keeping this manipulation item-specific. As a consequence, we assumed that the effect of ISb and CON_0 interaction might have been non-significant since Contingency (PRsb) included in the model explained all its variance. This was exactly the case: in all the control analyses, the removal of PRsb produced the same pattern of results, except for the interaction between ISb and CON_0, which became significant after this change. Thus, these consistent results confirmed our assumption that PRsb alone can explain all the variance of reactive control modulation and have several implications, as detailed in the following paragraph.

First, ISPC-induced reactive control and 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 Section 1.1), 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 contingency, were included in the statistical model. The third implication regards the fact that, although ISPC-induced reactive control alone was not significant, its interaction with proactive control to modulate the Stroop effect was significant even when contingency was in the model, suggesting that our experimental design was effective at yielding a reactive control effect partialled out from contingency, albeit an indirect one. This claim was further supported by another follow-up control analysis testing a model wherein PRsb and ISb were switched, so that, instead of ISb, PRsb was included in the three-way interaction to assess whether PRsb interacted with LWb in modulating the Stroop effect, that is, to exclude that the PRsb effect was masked by the inclusion of ISb. This, however, did not occur as the three-way interaction did not result significant, providing further evidence for the significant role of ISb in the three-way interaction and also indicating

that it was successfully partialled out from the effect of contingency (see Supplementary Materials at osf.io/aeH2d).

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 LWb and CON_0 was not only always significant in our main LMM and RCA analyses, but it was also accompanied by d values that were greater than .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 LWb 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 LWPC 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 LWPC is inevitably confused with ISPC since low-PC blocks are composed of low-PC items and high-PC blocks are composed of high-PC items (e.g., Blais et al., 2007; Blais & Bunge, 2010; Bugg et al., 2008). This is an inevitable consequence of LWPC manipulation. Indeed, although we tried to orthogonalize LWPC and ISPC as much as possible, our predictors were still correlated. However, by using multilevel modeling and including both predictors in the model, we were able to control for that confound. Therefore, since the LWPC by Congruency interaction consistently emerged, as opposed to the ISPC 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., 2023b). 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 proactive control independently operated by modulating the magnitude of the Stroop effect, even when reactive control was manipulated, whereas ISPC-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 ISPC 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 ISPC for congruent trials. However, although the correlation between our ISb and PRSb predictors was not that high (13% of shared variance), 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 ISPC 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 towards 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 ISPC-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 ISPC-induced reactive control effects. However, our results also revealed that the impact of reactive control still remained significant even after partialling 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 ISPC-induced reactive control as shown by one of our control analyses, as discussed above (also see Supplementary Materials at 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 ISPC-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 ISPC estimate is conditional to an intermediate level of the trial-level LWPC 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 a ISPC-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 ISPC-induced reactive control.

5.3. 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 LWPC estimates to exert proactive control, the stronger and more effective mode, made the implementation of reactive control unnecessary. Figure 5A clearly displays this, as can be seen from the pattern of observed results, when proactive control was high (blue line, hPro-hRea and hPro-lRea conditions), the Stroop effect did not decrease (and was even numerically larger) when both control strategies were available (hPro-hRea) as compared to when just proactive control strategies were implemented (hPro-lRea). By contrast, when proactive control was low (red line, lPro-hRea and lPro-lRea conditions), the Stroop effect was lower when reactive control was implemented (lPro-hRea), even if still higher than that observed under proactive control, as compared to when no control mode was available (lPro-lRea).

Importantly, the observed pattern of the interaction between proactive and reactive control in modulating the Stroop effects is well in line with the DMC 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 LWPC and ISPC, respectively. Indeed, DMC-inspired studies investigating or interpreting the behavioral and neurophysiological correlates of proactive and reactive control usually only varied LWPC (or LWPC-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 LWb was high (and thus proactive control was low), the Stroop effect decreased only when ISb was low (and thus reactive control was high), whereas it was very large when ISb was also high (and thus reactive control was low). In other words, high levels of LWPC 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 ISPC. 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 ISPC (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 (Tafuro et al., 2020; Bugg, 2017; Bugg & Crump, 2012).

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: i) the graphical representation displaying the model better showing the results provides a clearer and more easily interpretable overview of the pattern of results; ii) it encourages comparability with other works testing both similar and distinct hypotheses.

5.4. 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., 2023b), 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., 2023b), 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 intra-subject/inter-trial variance, has been shown to effectively provide more precise estimates of the effects of interest.

6. Conclusions

In two behavioral Experiments, we aimed to investigate the dual-mechanisms of control model (DMC) 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 LWPC and ISPC 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 LWPC-dependent proactive control mechanisms modulating Stroop performance regardless of confounders and also ISPC-dependent reactive control levels. Moreover, albeit we did not find evidence for the existence of specific ISPC-dependent reactive control effect, ISPC-induced reactive control still interacted with proactive control in modulating Stroop performance, with the characteristic pattern assumed by the DMC.

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; 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 LWPC and ISPC, we provided evidence for the interplay between proactive and reactive control in modulating the Stroop effect, which set the bases for further studies that deepen the understanding of the two control modes postulated by the DMC. Third, thanks to our analysis-level control approach used to control for (and estimate) the effect of contingency, we unveiled the interplay between ISPC-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 valuably contributed 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).
- Baayen, R., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3(2). <https://doi.org/10.21500/20112084.807>
- Blais, C., & Bunge, S. (2010). Behavioral and neural evidence for item-specific performance monitoring. *Journal of Cognitive Neuroscience*, 22(12).
- 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., & Egnér, T. (2019). Measuring adaptive control in conflict tasks. *Trends in cognitive sciences*. <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). <https://doi.org/10.1016/j.tics.2011.12.010>
- Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. *Variation in working memory*, 75, 106.
- 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, 567–587.
<https://doi.org/10.1037/a0035032>
- Bugg, J. M. (2017). Context, conflict, and control. *The Wiley handbook of cognitive control*, 79–96.
- 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, 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).
- 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). doi: 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).
- Chiew, K. S., & Braver, T. S. (2017). Context processing and cognitive control: From gating models to dual mechanisms. In *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). <https://doi.org/10.1017/S0140525X12000477>
- Cohen, J. D. (2017). Cognitive control: Core constructs and current considerations. In *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).
<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>

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, 14(4).

Gonthier, C., Braver, T. S., & Bugg, J. M. (2016). Dissociating proactive and reactive control in the Stroop task. *Memory & Cognition*, 44(5).

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).

<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).

<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).

Jacoby, L. L., Lindsay, D. S., & Hessels, S. (2003). Item-specific control of automatic processes: Stroop process dissociations. *Psychonomic Bulletin & Review*, 10(3).

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 G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*, 2 (pp. 743–777).

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).

<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).

<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).

<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). <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, 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). 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).

<https://doi.org/10.3758/CABN.7.1.1>

Parr, T., Holmes, E., Friston, K. J., & Pezzulo, G. (2023). Cognitive effort and active inference. *Neuropsychologia*, 184, 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). <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). 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). <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, 426–435.

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, 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). <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). <https://doi.org/10.1177/0098628316677643>

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, 18(6), Articolo 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, 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>

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>

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*, 1-28. <https://doi.org/10.3758/s13428-023-02215-0>

Viviani, G., Visalli, A., Finos, L., Vallesi, A., & Ambrosini, E. (2023b). A comparison between different variants of the spatial Stroop task: The influence of analytic flexibility on Stroop effect estimates and reliability. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-023-02091-8>