



Not All Stroop-Type Tasks Are Alike: Assessing the Impact of Stimulus Material, Task Design, and Cognitive Demand via Meta-analyses Across Neuroimaging Studies

Veronika I. Müller^{1,2} · Edna C. Cieslik^{1,2} · Linda Ficco^{3,4} · Sandra Tyralla⁵ · Amir Ali Sepehry⁶ · Taraneh Aziz-Safaie^{1,2} · Chunliang Feng⁷ · Simon B. Eickhoff^{1,2} · Robert Langner^{1,2}

Received: 25 January 2024 / Accepted: 29 July 2024
© The Author(s) 2024

Abstract

The Stroop effect is one of the most often studied examples of cognitive conflict processing. Over time, many variants of the classic Stroop task were used, including versions with different stimulus material, control conditions, presentation design, and combinations with additional cognitive demands. The neural and behavioral impact of this experimental variety, however, has never been systematically assessed. We used activation likelihood meta-analysis to summarize neuroimaging findings with Stroop-type tasks and to investigate whether involvement of the multiple-demand network (anterior insula, lateral frontal cortex, intraparietal sulcus, superior/inferior parietal lobules, midcingulate cortex, and pre-supplementary motor area) can be attributed to resolving some higher-order conflict that all of the tasks have in common, or if aspects that vary between task versions lead to specialization within this network. Across 133 neuroimaging experiments, incongruence processing in the color-word Stroop variant consistently recruited regions of the multiple-demand network, with modulation of spatial convergence by task variants. In addition, the neural patterns related to solving Stroop-like interference differed between versions of the task that use different stimulus material, with the only overlap between color-word, emotional picture-word, and other types of stimulus material in the posterior medial frontal cortex and right anterior insula. Follow-up analyses on behavior reported in these studies (in total 164 effect sizes) revealed only little impact of task variations on the mean effect size of reaction time. These results suggest qualitative processing differences among the family of Stroop variants, despite similar task difficulty levels, and should carefully be considered when planning or interpreting Stroop-type neuroimaging experiments.

Keywords Stroop · Cognitive control · Interference resolution · Neuroimaging meta-analysis · Activation likelihood estimation

✉ Veronika I. Müller
v.mueller@fz-juelich.de

¹ Institute of Neuroscience and Medicine, INM-7, Research Centre Jülich, Jülich, Germany

² Institute of Systems Neuroscience, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University, Düsseldorf, Germany

³ Department of General Psychology and Cognitive Neuroscience, Friedrich Schiller University, Jena, Germany

⁴ Department of Linguistics and Cultural Evolution, International Max Planck Research School for the Science of Human History, Jena, Germany

⁵ Institute for Experimental Psychology, Heinrich Heine University, Düsseldorf, Germany

⁶ Clinical Psychology Program, Adler University (Vancouver Campus), Vancouver, Canada

⁷ Key Laboratory of Brain, Cognition and Education Sciences, South China Normal University, Guangzhou, China

Interference Processing and the Stroop Paradigm

In everyday life, we are confronted with the need to regulate our behavior and override impulses and response tendencies that interfere with our overarching goals. Such interference often arises from the preferential processing of goal-irrelevant information because of habits, expectancy, attentional orientation, or memory-driven biases such as priming. Keeping behavior flexible and aligned with overarching goals requires cognitive control, which resolves interference by re-biasing the processing focus toward goal-relevant information (Diamond, 2013).

One of the best-known experimental tasks to elicit cognitive interference is the Stroop task. Introduced in 1935 by John Ridley Stroop (Stroop, 1935), the task requires participants to name the ink color of printed words naming semantically incongruent colors (e.g., the word “red” printed in blue ink) and of colored squares or crosses, which are used as a semantically non-interfering (“neutral”) control condition. Behaviorally, the Stroop interference effect manifests in healthy individuals as a delay in reaction time when naming the ink color of incongruent color-words, as compared to neutral conditions. Neuroimaging studies have shown that this effect goes along with increased activation in a widely distributed brain network including bilateral anterior insula, regions of lateral frontal gyrus and junction, intraparietal sulcus, and superior and inferior parietal lobules as well as the anterior midcingulate cortex and pre-supplementary motor area (Huang et al., 2020; MacLeod & MacDonald, 2000). These regions are often summarized as multiple-demand network (Duncan, 2010) and are generally recruited in tasks probing executive functioning.

Since its introduction, the Stroop task and Stroop-like phenomena have gained strong popularity in cognitive psychology and neuroscience as well as in clinical neuropsychological contexts. Stroop-like tests are widely adopted to assess executive impairments (Braga et al., 2022) and frontal lobe (dys)functioning (Cipolotti et al., 2016; Demakis, 2004), monitor drug effects (Pilli et al., 2013), and induce mental fatigue (Sun et al., 2021). Accordingly, the Stroop task is part of different psychometric test batteries, such as the Delis-Kaplan (Delis et al., 2001) or CANS-MCI (Tornatore et al., 2005) test battery, and different other test libraries (for example, PsyToolkit’s, <https://www.psytolkit.org/experiment-library/#exps>; or the Psychology Experiment Building Language Test Battery, <https://pebl.sourceforge.net>; Mueller & Piper, 2014). The Stroop test is especially widely used in the clinical context to assess cognitive impairments in conditions such as attention-deficit disorder (Lansbergen et al., 2007), cardiovascular

disease (Dintica et al., 2022; Shao et al., 2020), dementia and mild cognitive impairment (Rabi et al., 2020; Spieler et al., 1996), hepatic encephalopathy (Luo et al., 2020), obsessive-compulsive disorder (Gruner & Pittenger, 2017), or traumatic brain injury (Dimoska-Di Marco et al., 2011), to only name a few. A PubMed database query (on 11th of June, 2023) using the search string “stroop AND (clinic* OR patient* OR patholog* OR disease* OR disorder* OR syndrom* OR prodrom*)” yielded 5770 hits, further highlighting the outstanding relevance of this task in clinical and neuropsychological research.

Variations of the Task

Originally, the Stroop task was presented in separate blocks of trials per condition, in which performance was measured as the total amount of time needed to verbally name the colors of words in a list. Since Stroop introduced the task in 1935, different variants of the task have evolved (Lezak, 2012; Macleod, 1991), all aiming to capture how the processing of one stimulus dimension interferes with the processing of another one when two stimulus dimensions of a multidimensional stimulus overlap (Kornblum & Lee, 1995). In the behavioral literature, Stroop-like experiments differ in many aspects with potential impact on interference, including manipulations of the experimental design and type of stimulus presentations, semantic variations of the irrelevant dimension, manipulations of the probability of conditions, size of the stimulus and response set, response modality, and variations in the control condition but also in the stimulus material (Macleod, 1991). Additionally, studies have investigated the effects of the list-wise proportion of incongruent items (Lindsay & Jacoby, 1994), thus manipulating the amount of proactive and reactive control (Braver et al., 2008), as well as proportion effects on the item-specific level (item-specific proportion congruent effect; Jacoby et al., 2003). In neuroimaging, the most common variations pertain to the use of different control conditions, the way the different conditions are presented (presentation design), and the inclusion of additional cognitive demands as well as various types of stimulus material. Given that this study focuses on effects observed in neuroimaging experiments, only these variations that are common in neuroimaging will be further considered.

Experiments using Stroop-like tasks differ, for example, in the kind of *control condition* used for comparison with the incongruent target condition. While the original experiment employed color patches (and crosses for the investigation of practice effects), later studies used strings of different colored characters (like XXX, %%, \$\$\$, ***) or neutral (color-unrelated) words as control condition (for overviews, see Macleod, 1991, 2005). In addition, 35 years after

Stroop's initial experiment, a "facilitation" condition was introduced, in which target and irrelevant stimulus dimensions were made congruent to each other, like the word "red" printed in red (Dalrymple-Alford & Budayer, 1966). Subsequently, more and more experiments directly compared incongruent with congruent conditions. Especially in neuroimaging research, examining the Stroop effect via contrasting an incongruent with a congruent condition has become very popular. The main reason for this preference might be the fact that in both incongruent and congruent conditions, the same stimuli (e.g., words) are used, and only the congruency is manipulated (MacLeod, 2005), therefore controlling for possible stimulus effects.

Another factor of variation between Stroop tasks lies in the *presentation design*. While Stroop originally presented lists of stimuli per condition (i.e., blocked design), neuroimaging studies have typically used trial-by-trial stimulus presentations, where incongruent, neutral, and/or congruent conditions are either presented in blocks of trials (similar to the original version but with trial-wise, rather than list-wise stimulus presentations) or by mixing conditions.

Further, there are task variants that impose *additional cognitive demands*, by which an additional cognitive process is required for performing the task. For example, this applies (i) if the Stroop task is presented in the form of a match-to-sample task (i.e., requiring participants to match the color of a word to the meaning of another presented word or the color of a cue held in working memory) (e.g., Schulte et al., 2009; Zysset et al., 2001), (ii) if there is no fixed response mapping (i.e., response mappings for the different colors change from trial to trial), or (iii) if the Stroop task is combined with another task (e.g., a stop-signal task: Basten et al., 2011), and participants therefore have to constantly switch between task requirements.

On top of these variations of control conditions, design, and demand, also, the specific type of task or *stimulus material* greatly varies across studies, too. Besides color-word versions, studies on Stroop-type conflict processing also used auditory, pictorial, spatial, numerical, dimensional, shape, or emotional stimuli to induce interference (for review, see Macleod, 1991). The most common types of these other Stroop-like tasks used in neuroimaging experiments are the numerical, counting, spatial, and face-word Stroop task. In addition, there are affective versions that use the processing of emotional words to induce interference with color naming/counting (Gotlib & McCann, 1984; Whalen et al., 2006) or with the naming of the emotion expressed in another stimulus dimension (De Houwer & Hermans, 1994; Stenberg et al., 1998). While in the former version (naming the color in which an emotional word is printed), the emotional word meaning does not directly interfere with the response regarding the task (as the target dimension is color), in the latter version (naming the emotion expressed

in a face on which an emotional word is printed), the word meaning directly conflicts with the task-relevant dimension (identification of the emotional expression). Therefore, the emotional color-word interference task, which is traditionally called the emotional Stroop task, actually taps into a different phenomenon (Algom et al., 2004) with, importantly, no overlap between the different stimulus dimensions. This task is hence not within the focus of interest of the present work.

To conclude, even though all the aforementioned task variations are designed to investigate the interference of one stimulus dimension with another, they are still quite different. This might be due to adaptations of the task to the neuroimaging environment (e.g., Bush et al., 1998) and to the fact that different studies ask specific experimental questions. However, these differences in experimental setup and material might have an influence on the Stroop effect that is measured, both on the neuronal and also on the behavioral level.

Effects of Task Variations on Behavior and Its Neural Mechanisms

There is preliminary evidence for an impact of the aforementioned task features on behavior and brain activity related to interference processing in Stroop-type tasks, but only a few studies have directly compared different variations and results are often inconsistent.

Control Condition

Behavioral Level From some previous studies, there is evidence that the type of control condition is crucial for the size of the Stroop effect (Macleod, 1991; Parris et al., 2022). It has been suggested that the Stroop effect involves conflict at multiple levels and reflects a combination of response, semantic (subsumed as informational conflict), and task conflict (Augustinova et al., 2019). Depending on the specific type of incongruent, congruent, and neutral stimuli used, different amounts of conflict might be captured. In particular, congruent conditions do not induce informational conflict and usually lead to facilitation, i.e., shorter reaction times compared to neutral conditions. Therefore, the comparison of incongruent to congruent conditions is a confluence of facilitation benefits from congruent stimuli (that is not induced by neutral stimuli) and conflict costs from incongruent stimuli (Macleod, 1991). Thus, the nature of the Stroop effect (i.e., its constituent cognitive processes) differs depending on the type of control condition chosen. To complicate things, while congruent stimuli do not lead to informational conflict (conflict between the information provided by different stimulus dimensions), they do, however,

induce task conflict (Littman et al., 2019; MacLeod & MacDonald, 2000), as the two task sets of word reading and color naming are also in congruent conditions concurrently activated (Parris et al., 2022). Therefore, it is assumed that the comparisons against congruent conditions mainly reflect informational conflict, while those against neutral ones also involve task conflict (Shichel & Tzelgov, 2018).

Neural Level Regarding variations of the control condition, neuroimaging studies have not directly investigated differences from contrasting incongruent with congruent or neutral conditions, respectively. However, indirect evidence can be derived from studies that have shown that congruent compared to neutral conditions also activate some regions of the so-called multiple-demand system (Bench et al., 1993; Carter et al., 1995; Zysset et al., 2001), that is, regions involved in processing incongruent Stroop stimuli. This therefore indirectly implies that the choice of the control condition affects neuroimaging results.

Presentation Design

Behavioral Level Regarding task design, blocking or mixing conditions affects cognition and processing requirements. Reaction times are usually longer in mixed compared to blocked designs. These so-called mixing costs, however, can differ for different experimental conditions (Los, 1996). Specifically, for the Stroop task, studies have indicated that the way experimental conditions are varied (i.e., between individual trials or blocks of trials) leads to differences in interference and facilitation effects (Boucart et al., 1999; Salo et al., 2001). Salo et al. (2001), for example, found facilitation effects (of congruent trials) in blocked but not in mixed designs, as well as stronger interference effects in blocked compared to mixed designs. Hasshim and Parris (2018) also showed larger interference effects when blocking conditions and additionally demonstrated that presentation design primarily affects response conflict effects. However, in contrast to Salo et al. (2001) and Hasshim and Parris (2018), Floden et al. (2011) reported stronger interference effects for mixed than for blocked presentations.

Neural Level Only a few studies provide information on the neural effects of mixing or blocking experimental conditions in the Stroop task. Leung et al. (2000) indirectly compared the results of an event-related Stroop study with the results of a different study using a blocked design and found an overlap of 26% of voxels with more bilateral effects in the mixed presentation, whereas the blocked one revealed a more left-sided involvement. Additionally, sequence and adaptation effects have been observed in mixed designs (Egner, 2011; Egner & Hirsch, 2005), which suggest smaller interference effects when conditions are blocked. Furthermore, Floden

et al. (2011) found reduced activation of posterior medial frontal regions in a blocked compared to a mixed design. Finally, besides Stroop-specific effects, there are well-known general effects of the way experimental conditions are varied in neuroimaging: blocked designs usually have higher detection power (Birn et al., 2002; Clark, 2012), are easier to implement, and are easier for participants to perform; on the other hand, they are confounded by anticipation and adaptation (Clark, 2012).

Additional Cognitive Demands

Behavioral Level Furthermore, additional cognitive demands can increase the size of the Stroop interference effect. Penner et al. (2012), for example, reported larger reaction time differences between incongruent and congruent conditions in a matching task, where an additional forced-choice comparison was included (via an instruction to indicate if the color of a presented word was the same as the word meaning of another word), as compared to when there was no matching requirement.

Neural Level Similar to variations of the presentation design, there is only sparse evidence on the change of neural patterns with increasing cognitive demand in Stroop-type tasks. However, for cognitive-control tasks in general, it has been suggested that there is specialization in multiple-demand (MD) regions when demand is rather low (e.g., recruitment of left-sided regions for verbal tasks), but as cognitive load increases, more and more regions of the MD network are recruited (e.g., also recruitment of right-sided regions for verbal tasks) in a rather non-specific fashion (Shashidhara et al., 2019).

Stimulus Material

Behavioral Level Behaviorally, some studies have shown that the Stroop interference effect differs depending on the stimulus material used. For example, smaller Stroop effects for spatial compared to color-word versions have been reported (Banich, 2019; Capizzi et al., 2017; Hilbert et al., 2014) as well as stronger effects for emotional than non-emotional face-word versions (Chechko et al., 2013). In contrast, Mitchell (2005) and Zoccatelli et al. (2010) did not find differences in the Stroop interference effect between color-word, counting, and shape-word versions as well as between color-word and spatial versions.

Neural Level Mitchell (2005) used different stimulus materials in the fMRI scanner and reported stronger Stroop effects in the left dorsolateral prefrontal cortex (dlPFC) for a color-word compared to a counting Stroop task. Zoccatelli et al. (2010) found overlaps between spatial and color-word Stroop

versions in the anterior cingulate cortex (ACC), supplementary motor area (SMA), left inferior parietal lobe (IPL), right middle frontal gyrus, and cerebellum but in general stronger effects with larger activations for the color-word version. Banich et al. (2000) also compared spatial and color-word Stroop versions and found similar regions, but the location of activation varied between the different tasks. In line with behavioral findings, stronger Stroop effects in emotional compared to non-emotional task versions (Chechko et al., 2012, 2013) have been reported in inferior frontal gyrus, insula, SMA, cingulate cortex, and IPL, as well as visual and temporal areas.

Co-recruitment, Relative Specialization, and the Multiple-Demand Network

As mentioned above, solving Stroop-like conflict has been associated with the activation of a distributed fronto-parietal network (Cieslik et al., 2015; Huang et al., 2020) in numerous studies. This network is not only engaged during Stroop-like tasks but in general during a broad set of different tasks and is therefore commonly called multiple-demand network (MDN, Duncan, 2010, 2013). It has been suggested that this network creates mental control programs by combining the required components for the task (Duncan, 2013; Shashidhara et al., 2019). This is, on the one hand, reflected by co-recruitment of regions of the MDN, especially in conditions of high cognitive demand with an increase in MDN activation in more difficult compared to easy tasks (Duncan, 2010; Fedorenko et al., 2013; Shashidhara et al., 2019). Shashidhara et al. (2019), for example, demonstrated progressive recruitment of the entire MDN when adding complexity, time pressure, and reward to a simple spatial maze task, reflecting an increased need of integration in more demanding and motivating tasks. Additionally, relative specialization within the MDN has on the other hand also been demonstrated (Assem et al., 2022; Shashidhara et al., 2019), especially when demand is low, potentially reflecting specific aspects of the tasks (instruction, rule, stimuli). Previous meta-analyses comparing different tasks (Stroop, spatial interference, Go/No-Go, Stop-Signal, etc.) of cognitive action control (Cieslik et al., 2015) support this notion, but there is little information on relative specialization within one specific task.

Aim of the Current Study

In summary, Stroop-type interference phenomena are characterized by conflict that arises from two different stimulus dimensions (and their associated behavioral consequences) that show semantic overlap, leading to crosstalk and

enhanced processing costs. A plethora of studies have used this paradigm to investigate interference processing, and previous neuroimaging meta-analyses have identified regions of the MD network to be involved when processing Stroop-like tasks (Chen et al., 2018a; Cieslik et al., 2015; Huang et al., 2020; Song et al., 2017). However, beyond sharing the common feature of interference between two stimulus dimensions, there are still differences in some task features like control condition, presentation design, and stimulus material. How co-recruitment and relative specialization within the MD network are reflected during Stroop-like interference processing has not been investigated in detail and systematically. In particular, there is the question of whether involvement of the multiple-demand network can be attributed to the higher-order conflict that all of the task variations have in common, or if aspects that vary between task versions play a modulating role. Based on the fact that Stroop-like effects share the common property of an overlap between two dimensions of a multidimensional stimulus, one would expect only little variations in the brain regions recruited. Neuroimaging evidence is, however, inconclusive regarding such commonalities and also some differences were found. Thus, by using meta-analyses (activation likelihood estimation and robust variance estimation) across neuroimaging experiments, we aimed to investigate this issue systematically. Given that the task is frequently and routinely used in clinical and cognitive neuropsychology, it is crucial to elucidate which task version affects brain activation in which way, as differences in regional recruitment between versions of the task may, in turn, influence the sensitivity of the given version to detect a particular cognitive deficit.

Our first aim was to assess the impact of variations of Stroop-like tasks and experimental setup on the neural level (i.e., variation of spatial location of convergence). Second and as a follow-up analysis, we aimed to investigate the effects of the same factors on behavior reported in those studies (i.e., impact of variation on effect size). In particular, we asked the question if the variation of the following factors, which vary across neuroimaging settings, influences the size of the behavioral Stroop effect and the regions recruited during processing Stroop-like tasks: (1) stimulus material (e.g., color-words, emotional face-word combinations, or other stimuli), (2) type of control condition in the color-word Stroop task (i.e., contrasting the incongruent condition with neutral vs. congruent conditions), (3) presentation design (blocking or mixing of conditions) in the color-word Stroop task, and (4) imposing additional cognitive demand in the color-word Stroop task.

Given that all task variations share common properties of interference of two stimulus dimensions, we expected to find a consistent behavioral Stroop effect as well as core regions of the multiple-demand network recruited across all task variations. However, given previous findings, we additionally

expected some differences. With respect to the type of control condition chosen, we expected larger behavioral Stroop effects and regions of the MD system to be more consistently involved when contrasting incongruent with congruent, relative to contrasts with neutral control conditions, due to the added effects of facilitation and task conflict in congruent Stroop trials. Assuming that the presentation design would alter processing requirements, we also expected differences between blocked and mixed designs with respect to the average size of the Stroop effect and the spatial convergence of incongruency-related brain activity across studies. However, we have no specific hypothesis about the direction of the differences, as results in the literature are inconsistent, mixing costs are asymmetric, and mental and methodological effects operate differently. For experiments that impose additional cognitive demand during Stroop-type interference processing, we expected larger behavioral Stroop effects on average as well as a broader and more consistent recruitment of the MD network.

No study protocol was used.

Methods

Sample

As the primary goal of the current work was to investigate the neural correlates of the Stroop interference effect across various task implementations, literature search and experiment selection criteria were initially focused on neuroimaging studies. In a second step, all eligible experiments included in the neuroimaging meta-analysis were checked for eligibility for the behavioral meta-analyses. As a result, we also included studies in the neuroimaging meta-analysis that turned out to be not eligible for the behavioral meta-analysis (i.e., inclusion of studies that did not report all the necessary behavioral data for calculating effect sizes of behavioral effects). Thus, our behavioral results can only be generalized to the typical settings in neuroimaging experiments.

We did not seek to obtain an ethics vote for this study as our analyses did not include any individual participant data but were solely based on previously published aggregated data.

Literature Search and Inclusion and Exclusion Criteria

This study was part of a larger project on the evaluation of Activation Likelihood Estimation (ALE) meta-analysis. Here, we built on previous meta-analyses of supervisory control (Cieslik et al., 2015) as well as emotional interference processing (Chen et al., 2018a). To extend our sample

beyond the Stroop and Stroop-like experiments included in those previous studies, an additional literature search was carried out as part of three subprojects (not reported here), including experiments published before 30.05.2023. Three (LF, ST, TA) researchers conducted the literature search (conducted between 2020 and 2021) of experiments and extracted the relevant data (coordinates, space, sample size, and task variation for neuroimaging meta-analyses; reaction times, standard deviation/error, correlation, and task variation for behavioral meta-analysis). Another author (VM) double-checked the included experiments, extracted data, and updated the database in fall 2023 (which was again double-checked by a co-author). Disagreement in data extraction and inclusion was resolved by discussion between the respective two researchers who did the search and double-checked as well as by obtaining additional advice from other co-authors. Published neuroimaging experiments of Stroop tasks using functional magnetic resonance imaging (fMRI) or positron emission tomography (PET) were identified by queries of three databases: PubMed (<https://pubmed.ncbi.nlm.nih.gov/https://pubmed.ncbi.nlm.nih.gov/>), Google Scholar (<https://scholar.google.de>), and Web of Knowledge (<https://apps.webofknowledge.com>). The databases Scopus, the Networked Digital Library of Theses and Dissertations (NDLDT), and OADT.org were queried to identify pertinent theses, dissertations, and conference papers/posters.

In addition to the database queries, 96 authors were contacted and asked to provide result coordinates or images of contrasts of interest (see supplementary table S1 for information on studies for which additional information was provided) and/or information for effect size calculations if pertinent results were not reported in the publication (e.g., we asked authors of clinical studies for results of the healthy sample when only effects across patients and healthy controls were reported). Neuroimaging results provided by authors and shared as full images or coordinate tables were coded as three peaks per cluster to keep the number of peak coordinates similar to those usually reported in the literature. We aimed at a sample size as large as possible with a minimum of 17 experiments needed for calculating neuroimaging meta-analysis (Eickhoff et al., 2016; Frahm et al., 2022).

Due to the lack of a system for risk assessment of individual studies included in neuroimaging meta-analyses, quality assessment was implemented via the detailed coding of methodological features of each study and strict quality-related inclusion criteria, thereby excluding experiments not meeting those standards or not providing the relevant information for assessment.

In particular, the following inclusion/exclusion criteria were applied:

- We selected studies that investigated Stroop-like phenomena in a non-clinical adult sample and reported (or

- provided us upon request) coordinates of the results of contrasting the incongruent task condition with a congruent or neutral control condition.
- Only experiments were included that used two-dimensional stimuli with a task-relevant dimension that featured overlap with the irrelevant stimulus dimension, therefore using stimuli with a logical relationship between dimensions (Algom et al., 2022). We selected only those experiments where the two stimulus dimensions activate different processes (for example, *reading* versus *color naming*, *reading* versus *counting*), while experiments where the two stimulus dimensions activate the same process (e.g., global–local interference tasks; both task-irrelevant and task-relevant dimensions activate *letter reading*) were excluded.
 - Furthermore, to avoid strongly imbalanced heterogeneity on the level of presentation modality, only experiments using visual stimuli were considered, whereas experiments using other stimulus modalities or cross-modal settings were excluded.
 - Experiments with a Simon component were excluded, i.e., where stimulus position interferes with the response.
 - From the subset of emotional Stroop tasks, we excluded classic color-emotion-word and emotional counting Stroop tasks because they do not induce a semantic or response conflict (Feng et al., 2018) but induce a delay in response times due to attentional capture by the emotionality of the target word. Additionally, they don't meet the criteria of a logical relation between stimulus dimensions.
 - In addition, we only included experiments where a conflict was induced at both stimulus and response levels, as it is questionable if semantic conflict alone can be measured reliably (Hasshim & Parris, 2014, 2015). In turn, we excluded contrasts that only reflected one type of conflict, such as experiments where conflict only occurred at the level of the stimulus (perceptual/semantic conflict), which is, for example, the case when the response to a conflicting color-word is mapped onto the same response button (for example Chen et al., 2013) and is not part of the response set (for example Kim, 2010) or when semantically associated words are used instead of color-words (for example frog written in red ink, Banich et al., 2001). Furthermore, contrasts between response and semantic conflicts were also excluded (for example Chen et al., 2013).
 - Due to the limited number of experiments fulfilling all other inclusion criteria, we excluded spatial versions of the task ($n=2$) as well as studies posing additional cognitive demands for tasks other than the color-word Stroop ($n=3$).
 - Studies that re-analyzed data from the same participants as used in a different, already included experiment were excluded (for example, Videbech et al. (2003), who re-used the data from Ravnkilde et al. (2002)).
 - We only included studies that reported results of whole-brain group analyses as coordinates in a standard reference space (Talairach/Tournoux (TAL) or Montreal National Institute (MNI)). Thus, results obtained from region-of-interest analyses (ROI) were not considered. Reported coordinates resulting from neuroimaging analyses using the software SPM (Statistical Parametric Mapping) or FSL (FMRIB Software Library) were treated as MNI coordinates, unless a transformation into Talairach space was explicitly mentioned. Differences in coordinate space (MNI vs. Talairach space) were accounted for by transforming coordinates reported in Talairach space into MNI coordinates using a linear transformation (Lancaster et al., 2007).
 - Results from studies with patients or children were excluded as were experiments reporting between-group effects (for example, age- or disease-related effects) or pharmacological interventions. However, we did include clinical or intervention studies that reported within-group effects separately for the (healthy adult) control group or effects at baseline, respectively.
- These criteria led to the inclusion of 115 studies reporting 133 experiments (see supplementary table S3 for the checklist for neuroimaging meta-analyses and Fig. 1 for an illustration of the workflow of the current study). Of those studies, 77 studies from 68 independent labs reported (or provided) sufficient behavioral data used for calculating the effect size for the meta-analysis of the behavioral Stroop effect (see method description of the follow-up analysis for further details). Tables illustrating the data separately for the neuroimaging and effect size meta-analyses are presented in the supplement (table S1 and S2).

Neuroimaging Meta-analyses

Coding of Experiments

Control Condition of Color-Word Stroop Tasks Each color-word Stroop experiment was classified according to the condition to which the incongruent condition was compared (i.e., congruent or neutral). For either type of control condition, a separate meta-analysis was calculated: (1) congruent control ($I > C$, 42 experiments) and (2) neutral control ($I > N$, 34 experiments). Experiments that imposed additional cognitive demand were excluded from these analyses, as were experiments using different stimulus materials (i.e., other task types than the color-word Stroop version). Two supplementary meta-analyses were performed for which experiments reporting contrasts against a neutral control condition

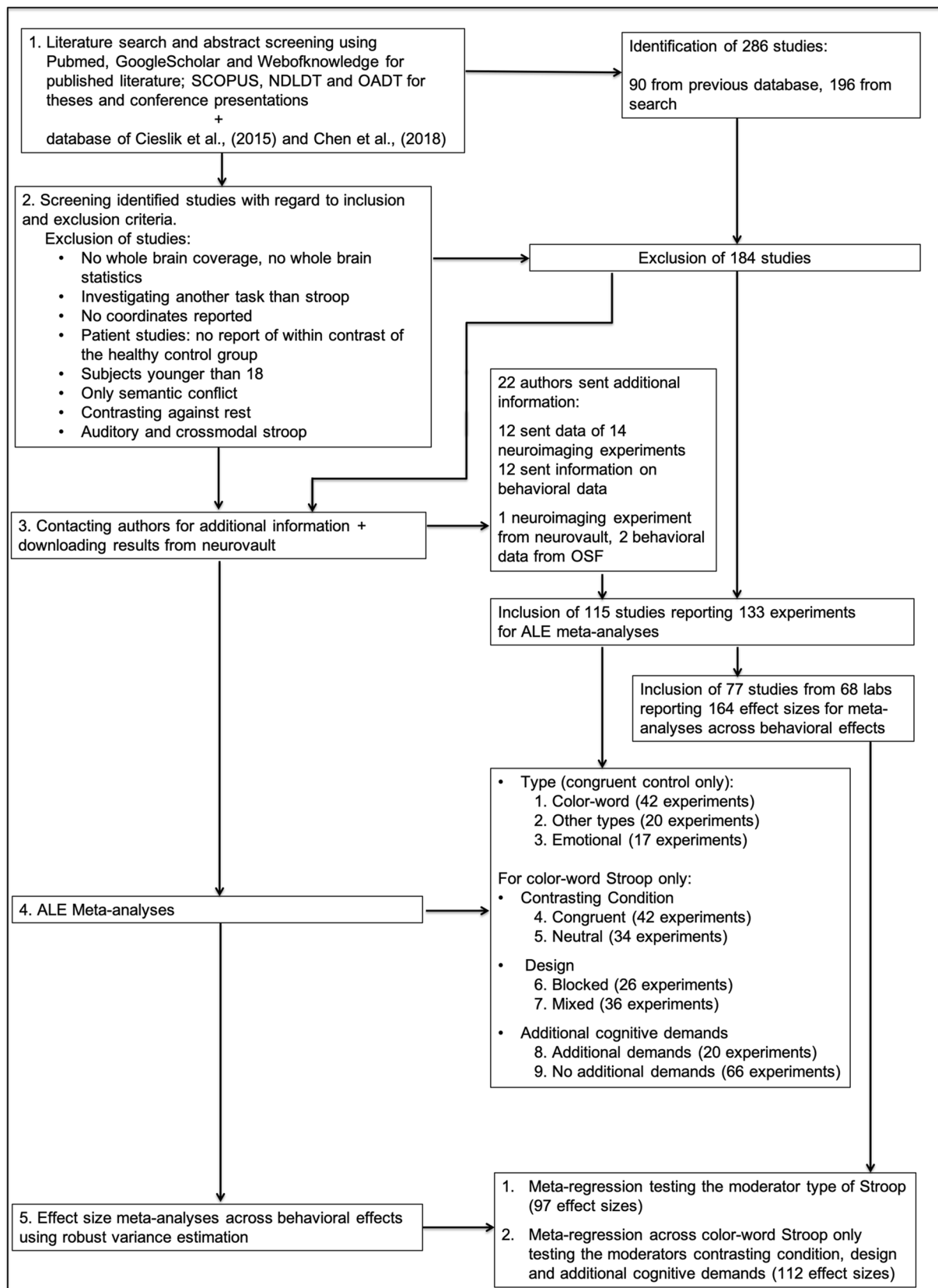


Fig. 1 Illustration of the workflow of the meta-analytical study

were separated into those using neutral words (18 experiments) and those using symbols/letters (17 experiments).

Presentation Design of Color-Word Stroop Tasks Color-word Stroop experiments were further classified according to the way the experimental conditions were varied across trials as either blocked (trials of the same condition presented in blocks of trials) or mixed designs (random or pseudorandom mix of conditions across individual trials). We only classified experiments as a mixed design if also data analysis was done in an event-related manner. That is, experiments using a mixed presentation within blocks (i.e., congruent and incongruent blocks interspersed with neutral trials), which then compared different blocks with each other, were excluded from this analysis of design effects. For either type of design, a separate meta-analysis was calculated: (1) blocked (26 experiments) and (2) mixed designs (36 experiments). Again, experiments that imposed additional cognitive demand were excluded from these analyses, as were experiments using different stimulus materials. Studies reporting eligible results from the same participants in more than one experiment (e.g., studies reporting both incongruent > congruent and incongruent > neutral contrasts) were coded as one experiment to avoid biasing the meta-analysis by non-independent data (Müller et al., 2018a).

Additional Cognitive Demand in Color-Word Stroop Tasks Another meta-analysis was calculated across color-word Stroop experiments that imposed some additional cognitive demand on top of the Stroop-type interference processing (20 experiments). These experiments included matching paradigms (for example, Zysset et al., 2001, introducing an additional forced-choice comparison where participants had to indicate if the color of a presented word is the same as the word meaning of another presented word) as well as combinations with different tasks (e.g., combined stop-signal and Stroop tasks). This meta-analysis was compared to an analysis across all color-word experiments without additional cognitive demands (across both kinds of control conditions and design, 66 experiments). Experiments using different stimulus materials were excluded, and studies reporting eligible results from the same participants in more than one experiment were treated as one experiment (cf. above).

Type of Stimulus Material Finally, we performed a meta-analysis of other forms of Stroop tasks, that is, those that used other than color-word stimuli to induce Stroop-type interference. We categorized these experiments as follows: counting Stroop, numerical Stroop, face Stroop, spatial Stroop, or emotional face Stroop. We created three pools of experiments for separate meta-analyses, (1) emotional Stroop (17 experiments), (2) color-word Stroop (42

experiments), and (3) other versions (20 experiments), which were grouped together due to the limited number of experiments per stimulus material (counting, 8; numerical/size, 6; face, 6). Importantly, we only included experiments that used a congruent condition as control because none of the emotional face Stroop experiments and only very few of the other types of Stroop used a neutral control condition. Furthermore, we excluded experiments that imposed additional cognitive demand.

Please note that not all of the experiments could be classified with regard to all moderators of interest, and therefore, some experiments were included only in some but not all analyses (supplementary tables S1 and S2 show the coding of each experiment).

Activation Likelihood Estimation Algorithm

Standard analysis procedures were performed as used in previous ALE studies (cf. Caspers et al., 2010; Chase et al., 2015; Cieslik et al., 2015; Langner & Eickhoff, 2013). In brief, coordinate-based meta-analyses were performed to identify consistent activations across experiments by using the revised Activation Likelihood Estimation algorithm (Eickhoff et al., 2009, 2012; Turkeltaub et al., 2012) implemented as in-house Matrix Laboratory (MATLAB, 2019) tools. This algorithm aims to identify areas showing convergence of reported coordinates across experiments that are higher than expected for a random spatial association. Reported foci are not treated as single points but rather as centers for 3D Gaussian probability distributions capturing the spatial uncertainty associated with each focus. The width of these uncertainty functions is determined by the between-subject (uncertainty of spatial localizations between different subjects) and between-template (uncertainty of spatial localizations between different spatial normalization strategies) variance, which represents the main components of this uncertainty. Importantly, the between-subject variance is weighted by the number of participants per experiment, accommodating the notion that larger sample sizes should provide more precise approximations of the true activation effect and should therefore be modeled by smaller Gaussian distributions (Eickhoff et al., 2009). The probabilities of all foci reported in a given experiment were then aggregated for each voxel, resulting in a modeled activation (MA) map for that experiment (Turkeltaub et al., 2012). To ensure that results were not driven by studies reporting more than one eligible contrast obtained in the same group of participants (e.g., a study reporting both $I > C$ and $I > N$ contrasts), different contrasts included in one meta-analysis were coded as one experiment. Taking the union across the MA maps yielded voxel-wise ALE scores describing the convergence of results at each voxel of the brain. To distinguish true

convergence across experiments from random overlap, ALE scores were compared to a null distribution (analytically derived, see Eickhoff et al., 2012) that reflects a random spatial association between experiments. Conceptually, the null distribution can be formulated as sampling a voxel at random from each of the MA maps and taking the union of these values in the same manner as done for the (spatially contingent) voxels in the true analysis. The p -value of an ALE score was then given by the proportion of equal or higher values obtained under the null distribution. The resulting non-parametric p -values for each meta-analysis were then thresholded at a cluster-level corrected threshold of $p < 0.05$ (cluster-forming threshold at voxel level, $p < 0.001$). Cluster-level family-wise error (FWE) correction was performed as suggested by Eickhoff et al. (2016) and Frahm et al. (2022) and described in detail in previous meta-analyses (Bzdok et al., 2012; Rottschy et al., 2012). First, the statistical image of the uncorrected voxel-wise p -values of the original analysis was thresholded at the cluster-forming threshold of $p < 0.001$. Then, the size of the clusters surviving this threshold was compared against a null distribution of cluster sizes. This null distribution of cluster sizes was derived by simulating 10,000 datasets of randomly distributed foci with identical properties (number of foci, uncertainty) as the original dataset. This distribution was then used to identify the cluster size that was only exceeded in 5% of all random simulations.

Individual Experiment Contributions All clusters of significant convergence across experiments were further analyzed with regard to which experiments actually contributed to convergence. This was done by testing how much each included experiment contributed to the summarized ALE value. In particular, for each cluster and each experiment, the summarized ALE value of all voxels of the cluster with and without the experiment in question was calculated. If the summarized ALE value across all voxels of a cluster decreased when removing an experiment, that experiment was counted as contributing to the convergence of that cluster.

Contrasts and Conjunctions To determine those voxels where a significant effect was present in two separate analyses, conjunctions were computed using the conservative minimum statistic (Nichols et al., 2005). That is, only regions significant on a corrected level in each individual ALE analysis were considered. To exclude smaller regions of presumably incidental overlap between the thresholded ALE maps of the individual analyses, an additional extent threshold of 25 voxels was applied (Langner & Eickhoff, 2013; Müller et al., 2018a, 2018b).

Differences between conditions were tested by computing contrast analyses as used in previous studies (cf. Cieslik

et al., 2015; Rottschy et al., 2012). In particular, the difference between two ALEs was compared to a random distribution of differences under the null hypothesis of label exchangeability. First, the real difference between the two individual analyses was determined by computing the voxel-wise difference between the unthresholded ALE maps of each analysis (Eickhoff et al., 2011). Second, we determined a null distribution of differences. This was done by pooling all experiments contributing to either analysis and randomly dividing them into two groups of the same size as the two original sets of experiments. ALE scores for these two randomly assembled groups were calculated, and the difference between these ALE scores was recorded for each voxel in the brain. Repeating this process 25,000 times then yielded an expected distribution of ALE score differences under the assumption of label exchangeability. The observed difference in ALE scores was then tested against this null distribution yielding a posterior probability that the true difference was not due to random noise in an exchangeable set of labels, based on the proportion of lower differences in the random exchange. The resulting probability values were thresholded at $p > 0.95$ (95% chance for true difference) and inclusively masked by the respective main effects, i.e., the significant effects of the ALE analysis for the particular condition. In addition, an extent threshold of 25 voxels was applied.

Follow-up Analysis: Meta-analyses Across Effect Sizes of the Behavioral Stroop Effect

Two authors screened all 115 studies included in the neuroimaging meta-analysis for eligibility for the behavioral meta-analysis. We only focused on the Stroop effect on response speed, as only some studies report data on accuracy. Studies were included if they reported sample size as well as mean reaction time (RT) and standard deviation or standard error of RT for the incongruent and congruent and/or neutral conditions. Numerical results presented in figures only were extracted with the web-based application of PlotDigitizer (Rohatki, 2021). Additionally, authors were contacted and asked to provide data and/or further information (see supplementary table S2 for information on which studies provided additional information). We included all available information in our analyses and thus allowed for multiple experiments (effect sizes) per study (see the “Robust Variance Estimation” section for detailed information on how we treated correlated effects). Disagreement on extracted data between authors was solved by discussion and with additional advice from co-authors. In total, 77 studies from 68 independent labs (studyIDs) reporting 164 experiments were included in the effect size meta-analyses of the behavioral Stroop effect.

The pool of data of the follow-up analysis features two different kinds of dependencies that need to be considered for the meta-analysis. First, the Stroop effect is generally obtained from within-subject designs, which requires considering the correlation between conditions (incongruent and congruent/neutral) for calculating the estimates for the effect size meta-analysis (Borenstein et al., 2009). Second, most studies reported multiple experiments per study leading to correlated effects within studies. To account for the former, we estimated the correlation between conditions (see detailed description below) by imputing a correlation coefficient from experiments where this information was available. The dependency structure resulting from including multiple experiments per study, in turn, was accounted for by robust variance estimation (see detailed description below).

Estimation of the Correlation Between the Incongruent and Congruent/Neutral Condition

For calculating effect sizes (ES) and standard errors (SE) for repeated measurements as obtained from within-subject designs typically employed in experiments on Stroop interference, not only mean RT and standard deviation (SD)/SE of the different conditions are necessary but also the correlation between RT scores of either condition. This correlation is, unfortunately, rarely reported. For experiments that provided a *t*-statistic between conditions, the standard deviation of differences, or figures with single-subject data, the Pearson correlation coefficient could be calculated from these values. In other cases, authors were contacted for information on the correlation coefficient or data from which the correlation could be derived. Of the 164 effect sizes included in total, we could obtain a correlation coefficient (or authors provided it on request) for 79 (of 35 studies) of them. From these 79 coefficients, we imputed a Pearson correlation coefficient by calculating an aggregated mean correlation coefficient using robust variance estimation (RVE) meta-analysis (see description of the method of RVE below), which was then applied for the remaining experiments where no correlation was available. The imputed correlation coefficient across all available correlations was 0.91. In order to test the impact of *r* on the estimates of the meta-analysis, a sensitivity analysis was performed. This was done by first calculating an intercept-only RVE random-effects meta-analysis across all experiments for which the real correlation was available and then repeating the same analysis but replacing the real correlation coefficient with a fixed one. In total, ten analyses were calculated (real *r*, *r* from 0 to 0.9 in steps of 0.1). The results of the sensitivity analysis revealed that estimates of the meta-analysis varied with changing correlation coefficients (table S4a). We therefore ran all behavioral meta-analyses three times: (1) using the observed correlation where available and the imputed correlation (0.91) for all

experiments with a missing *r*, (2) using the observed correlation where available and a plausible minimum correlation coefficient for all other experiments (assuming a minimum coefficient of 0.6 given that for RT in similar tasks intercorrelations of $r > 0.6$ have been reported (see, e.g., Jensen & Reed, 1990; Larson et al., 1988), and (3) using the imputed correlation coefficient (0.91) for all experiments to keep this factor constant across experiments. Only the results of the first analysis approach will be presented, as there were no differences in results between the three approaches.

Calculation of Standardized Effect Sizes and Variance

For each experiment, standardized effect sizes and standard errors were calculated in R 4.1.1. (R Core Team, 2021) using the formula of Borenstein et al. (2009) for repeated measurements and correcting for small sample bias (Hedges *g* and SE_g; Borenstein et al. (2009)). In detail, ES and SE were calculated based on RT means and SD/SE of the conditions as well as their corresponding Pearson correlation coefficient (*r*) for the 79 experiments where *r* was available; for all other experiments, ES and SE were based on the reported RT means and SD/SE as well as the imputed correlation coefficient of 0.91 (as described above). Effect sizes were calculated such that positive values reflect longer reaction times in the incongruent compared to the respective control condition (i.e., interference costs).

Robust Variance Estimation

As the inclusion of multiple effect sizes per study violates the assumption of independence of ES (Lipsey & Wilson, 2001), we adjusted our analysis for ES dependency by using robust variance estimation (RVE; Hedges et al., 2010; Tipton, 2013, 2015) using the *robumeta* package (version 2.0) in R (Fisher et al., 2017). Meta-analyses in general aggregate effect sizes by giving stronger weights to ES values with higher precision by inverse-variance weighting. In RVE, inverse-variance is also used, but additionally, the dependency structure between ES values is estimated and the weights adjusted accordingly (Hedges et al., 2010; Tanner-Smith & Tipton, 2016; Tipton, 2013). Based on the recommendation of Tanner-Smith and Tipton (2014), to determine the weighting scheme according to the most prevalent dependency structure, we used the correlated effect weighting scheme, as the highest amount of dependency resulted from having multiple effect sizes per study (only for ten studies, dependencies arose from hierarchical effects). For rho, we assumed a default of $\rho = 0.8$ as the within-study effect size correlation (Fisher & Tipton, 2015), necessary for RVE, as sensitivity analyses with varying ρ confirmed that estimates are not affected by ρ (table S4b).

Moderators of Interest and Meta-regression Models

An intercept-only random-effects RVE model across all effect sizes of the incongruence effect in reaction time was calculated for estimating an aggregated effect size across all experiments. As heterogeneity parameters (i.e., I^2) estimated in RVE are not precise estimates of variance parameter estimates (see Tanner-Smith & Tipton, 2016), we did not test for heterogeneity. Then, two RVE mixed-effects meta-regression models were estimated, one testing different moderators for color-word Stroop tasks on the incongruence effect simultaneously and another one testing the single moderator “stimulus material.” For the first model, the moderators (i) control condition (congruent vs. neutral), (ii) presentation design (mixed vs. blocked presentation of conditions), and (iii) additional cognitive demand (yes vs. no) were included. In the second meta-regression model, only the moderator stimulus material (color-word/emotional/other) was included. Similar to what we did in the neuroimaging meta-analyses (see the “Coding of Experiments” section), we only included contrasts against congruent conditions in the second model as there were no contrasts against neutral control conditions for emotional Stroop and only very few for other types. Wald tests implemented in the clubSandwich package version 0.5.3 in R (Pustejovsky, 2021) were used as omnibus tests for categorical moderators with more than two categories.

Additional explorative analyses modeling the moderators type of neutral control condition, interaction between design and control condition, response modality, and a number of different colors can be found in the supplement (tables S5 and S6).

Analyses of Sample Bias, Outliers, and Robustness of Results

Sampling bias was examined by calculating an RVE meta-regression model with the standard error of Hedges g effect size (SE g) as a moderator (Rodgers & Pustejovsky, 2020), which is similar to Egger’s regression but adjusted for dependent effect sizes. The significance of the moderator was taken as an indicator of funnel plot asymmetry and therefore some sort of bias.

Additionally, influential ES values were identified as outliers by using case deletion diagnostics (Viechtbauer & Cheung, 2010) for both meta-regression models described above. Here, we fitted random-effects meta-analyses ignoring the dependency of effect sizes by using the influence function in the metafor package (3.4–0) in R (Viechtbauer, 2010).

Transparency and Openness

We adhered to the Journal Article Reporting Standards for Quantitative Research in Psychology (Table 9 from Appelbaum et al., 2018) and the guidelines for neuroimaging meta-analyses (Müller et al., 2018a). We report how we determined our sample size, all data exclusion criteria, all manipulations, and all measures in the study. Analysis code of neuroimaging and behavioral meta-analyses can be found on the open science framework (OSF; https://osf.io/dt3kj/?view_only=995297bb53574583b1a0dda978f7f341); result files of the ALE meta-analysis are additionally available at ANIMA (<https://anima.fz-juelich.de/>; Reid et al., 2016). Behavioral data was analyzed using R, version R 4.1.1. (R Core Team, 2021); RStudio (RStudio Team, 2021); and the packages robumeta, version 2.0 (Fisher et al., 2017), metafor, version 3.4–0 (Viechtbauer, 2010), and clubSandwich, version 0.5.3 (Pustejovsky, 2021). Neuroimaging meta-analyses were analyzed with in-house Matlab, version 9.7.0.1471314 (MATLAB, 2019) tools. This meta-analytical project was not preregistered.

Results

Description of Included Experiments

We included 115 studies in total, reporting 133 experiments of neuroimaging results and 164 experiments of behavioral effects. The studies included a mean number of 29 participants with a mean age of 30 and on average an equal ratio of males and females. Eighty-five percent of the neuroimaging experiments required a manual response, and incongruent trials were presented with a probability of 42% on average. Seventy-two percent of the experiments used color-word stimuli. Of these color-word experiments, 57% reported contrasts against a congruent control condition, 41% presented conditions in blocks, and 21% included an additional component of demand. Experiments used on average 3.8 different colors as stimuli and 3.3 response alternatives (Table 1).

Neuroimaging Meta-analyses

Table 2, part B, provides an overview of all analyses and the results. For the meta-analyses across neuroimaging results, we first investigated the effects of control condition, presentation design, and additional cognitive demand across color-word Stroop experiments and then convergence across the different types of stimulus material used.

Table 1 Summary of included experiments describing the specific characteristics of experimental setups and participants

	All	Color-word only
% experiments using blocked design	34	41
% experiments using a congruent control	69	57
% experiment with no additional demand	85	79
% experiments color-word stimulus material	72	-
% experiments with manual response	85	80
Mean probability (SD) of incongruent trials	41.7 (10.6)	40.5 (11)
Mean (SD) number of participants	28.9 (26.8)	26.2 (21.5)
Mean (SD) age	30.1 (10.5)	29.8 (10.8)
Mean % (SD) of females	49.3 (23.1)	46.5 (24)
Mean (SD) number of response possibilities	3.1 (0.9)	3.3 (0.9)
Mean (SD) number of colors	-	3.8 (0.8)

Color-Word Stroop Task

Control Condition Two meta-analyses were calculated, one across all color-word experiments that contrasted against a congruent ($I > C$) and one across experiments with a neutral control condition ($I > N$). The two meta-analyses revealed similar regions of convergence (Fig. 2A and B, supplementary table S7 for main effects). Conjunction analysis, testing for regions that are significant in both meta-analyses, revealed convergence of both analyses in bilateral inferior/middle frontal gyrus/junction and anterior insula (aINS), left intraparietal sulcus (IPS), and posteromedial frontal cortex (pmFC; see Fig. 2C, shown in yellow for the conjunction).

To test where the convergence of results of the two meta-analyses significantly differed, a contrast analysis was calculated. This contrast between the two sets of results revealed stronger convergence for $I > C$ in the left inferior frontal junction and the pre-supplementary motor area (pre-SMA) (Fig. 4C shown in red) and for $I > N$ in the left inferior parietal lobe and posterior IPS, left middle frontal gyrus, and bilateral aINS (Fig. 2C, shown in green).

Results of supplementary analyses for which $I > N$ experiments were separated depending on whether they used words or non-words (letters or symbols) as neutral control conditions are shown in supplementary Figure S1. Results are similar to the effects found for the combined analysis, with the only difference that convergence in the left aINS seems to be mainly driven by experiments using symbols and letters, while convergence in the right inferior/middle frontal gyrus/junction was only found in the analysis across experiments using neutral words.

In summary, the use of congruent but also neutral control conditions consistently involved regions of the multiple-demand system (bilateral aINS, dorsolateral prefrontal cortex, left IPS, and pmFC), with a slightly larger extent of most clusters when contrasting against a neutral control condition.

Blocked Versus Mixed Presentation of Experimental Conditions Two meta-analyses were calculated: one across experiments that presented the different Stroop task conditions in blocks, and one across experiments implementing a mixed design. Figure 3A and B and supplementary table S8 present the main effects of blocked and mixed presentation designs, respectively. The conjunction between the meta-analyses for either type of presentation design revealed common convergence in the left inferior frontal gyrus (IFG), left IPS, and left aINS as well as pmFC (Fig. 3C, shown in yellow for the conjunction).

Contrast analyses that directly compared the results between both meta-analyses showed stronger convergence in more anterior pmFC and right orbitofrontal cortex for color-word Stroop experiments with a blocked design (Fig. 3C in red), whereas pre-SMA, bilateral aINS/frontal operculum, left middle intraparietal sulcus (mIPS), left fusiform gyrus, left precentral gyrus, and bilateral middle frontal gyrus were more consistently found for experiments using a mixed design (Fig. 3C in green).

In summary, our analyses of the impact of presentation design on the neural correlates of the Stroop effect point to a mainly left-sided convergence of interference-related brain activity in blocked designs and stronger and more bilateral convergence when task conditions are mixed. Additionally, a differentiation within the pmFC was found, with stronger convergence for mixed presentation designs in pre-SMA, more consistent activation for blocked designs in a more anterior cluster, and common convergence for both types of design located between these two clusters.

Additional Cognitive Demand Two meta-analyses were calculated to investigate the effect of additional cognitive demands on convergence; one analysis was calculated across experiments where an additional process was required for performing the task (matching, switching between task requirements, or response mapping) and one across experiments without an additional demand. Figure 4A and B and supplementary table S9 present the results of the main effects of the two meta-analyses of experiments with or without additional cognitive demand. The conjunction across both revealed common convergence for both analyses in bilateral aINS, bilateral middle/inferior frontal gyrus, bilateral IPS, and pmFC (Fig. 4C denoted in yellow).

Contrast analyses revealed stronger convergence for experiments with additional cognitive demand in left aINS, bilateral inferior frontal junction (IFJ), and right middle

Table 2 Overview of all analyses and the main results

Analysis	Main results
<i>A. Meta-analyses of behavioral effects</i>	
1. Color-word Stroop only testing the moderators control condition, design, and additional cognitive demands	No significant effects for any moderator
2. Stroop and Stroop-like phenomena testing the moderator stimulus material	Significant effect of stimulus material: Larger effects for color-word Stroop compared to emotional Stroop Larger effects for color-word Stroop compared to other types of Stroop (when removing outliers)
<i>B. Neuroimaging meta-analyses</i>	
<i>Color-word Stroop:</i>	
Control condition	
1. Congruent control	Convergence in mostly bilateral multiple-demand system (bilateral anterior insula, lateral prefrontal cortex (PFC), left intraparietal sulcus (IPS), left dorsal premotor cortex, and posterior medial frontal cortex (pmFC))
2. Neutral control	Convergence in bilateral multiple-demand system (bilateral anterior insula, lateral PFC and IPS, pmFC)
Contrast analysis control condition	Slightly larger clusters for neutral compared to congruent control condition
Design	
3. Blocked	Convergence in left-sided and medial multiple-demand regions (anterior insula, lateral PFC, IPS, pmFC) and right orbitofrontal cortex
4. Mixed	Convergence in bilateral multiple-demand system (bilateral anterior insula, lateral PFC and IPS, pmFC, left dorsal premotor cortex (dPMC)) and left fusiform gyrus
Contrast analysis design	Stronger and more bilateral convergence in most multiple-demand regions as well as fusiform gyrus when conditions are mixed Stronger convergence in the right orbitofrontal cortex for blocked designs Differentiation within pmFC with stronger convergence for mixed presentation in pre-supplementary motor area (pre-SMA), more consistent activation for blocked designs in a more anterior cluster and convergence found for both analyses located between those two clusters
Additional cognitive demands	
5. Additional demand	Convergence in bilateral multiple-demand system (bilateral anterior insula, lateral PFC and IPS, pmFC)
6. No additional demand	Convergence in bilateral multiple-demand system (bilateral anterior insula, lateral PFC and IPS, pmFC)
<i>Stroop and Stroop-like phenomena:</i>	
Stimulus material	
7. Color-word Stroop	Convergence in mostly bilateral multiple-demand system (bilateral anterior insula, lateral PFC, left IPS, left dorsal premotor cortex, and pmFC)
8. Emotional Stroop	Convergence in regions of the multiple-demand system (bilateral lateral prefrontal cortex, left IPS, left ventral dorsal premotor cortex, right anterior insula, pmFC)
9. Other types	Convergence in pmFC, right anterior insula, and left anterior IPS
Conjunction across all stimulus material	Consistent recruitment of pmFC (pre-SMA and aMCC) and right anterior insula
Contrast analyses	Specific association of emotional Stroop with ventral dPMC (stronger convergence compared to color-word and other types) Specific association of color-word Stroop with lateral prefrontal cortex (stronger convergence compared to emotional and other types) other types of Stroop showed in general less convergence and only compared to color-word version stronger convergence in aIPS

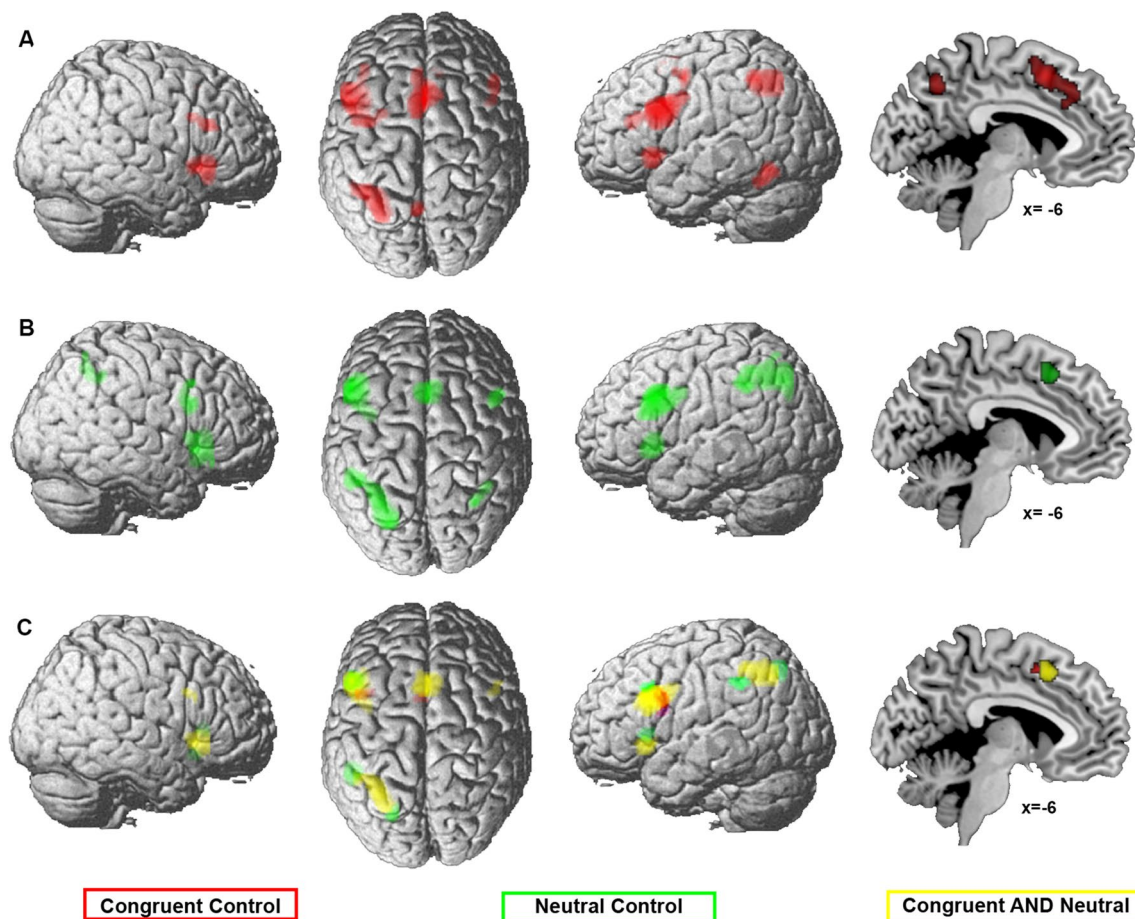


Fig. 2 Meta-analysis across experiments using congruent or neutral conditions as control. This figure illustrates the results of the meta-analyses across color-word Stroop experiments comparing incongruent with **A** congruent (42 experiments) or **B** neutral control (34 experiments) conditions as well as the conjunction (**C** in yellow)

and contrasts (**C**, red, stronger convergence for $I > C$; green, stronger convergence for $I > N$) between both analyses. The color intensity of the renderings (the first three columns) reflects the distance from the brain surface (lower intensity denoting further distance from the surface)

frontal gyrus (Fig. 4C in red) and for experiments without additional cognitive demand in the right orbitofrontal cortex, left middle frontal gyrus, and dorsal premotor cortex (dPMC, Fig. 4C denoted in green).

In summary, additional cognitive demand largely recruited the same network as without demand.

Different Types of Stimulus Material

We calculated three meta-analyses to investigate the influence of stimulus material: one meta-analysis across color-word Stroop experiments, one across emotional picture-word experiments, and one across other types of Stroop (including numerical, counting, and non-emotional picture-word Stroop varieties). The results of the three meta-analyses are shown in Fig. 5 and supplementary table S10. The convergences observed in the three meta-analyses were then compared to

each other by a three-way conjunction as well as conjunctions and contrasts between all pairs of stimulus material.

All three meta-analyses (revealed via conjunction) showed common convergence only in pmFC (pre-SMA and anterior midcingulate cortex) and right aINS (Fig. 6 in pink).

The conjunction analysis across color-word and emotional Stroop tasks revealed that both recruit bilateral inferior/middle frontal gyrus, right aINS, left mIPS, dorsal premotor cortex (dPMC), and pmFC (Fig. 6 in yellow). Emotional Stroop showed stronger convergence in the left ventral dorsal premotor cortex (dPMC) and right aINS (Fig. 7A in red), while color-word revealed more consistent recruitment of left middle frontal gyrus, pre-SMA, and left mIPS (hIPS1, Fig. 7A in green).

The conjunction analysis across color-word and other types (Fig. 6 in cyan) of Stroop did not reveal any additional overlap beyond that found in the three-way conjunction. For color-word Stroop tasks, stronger convergence

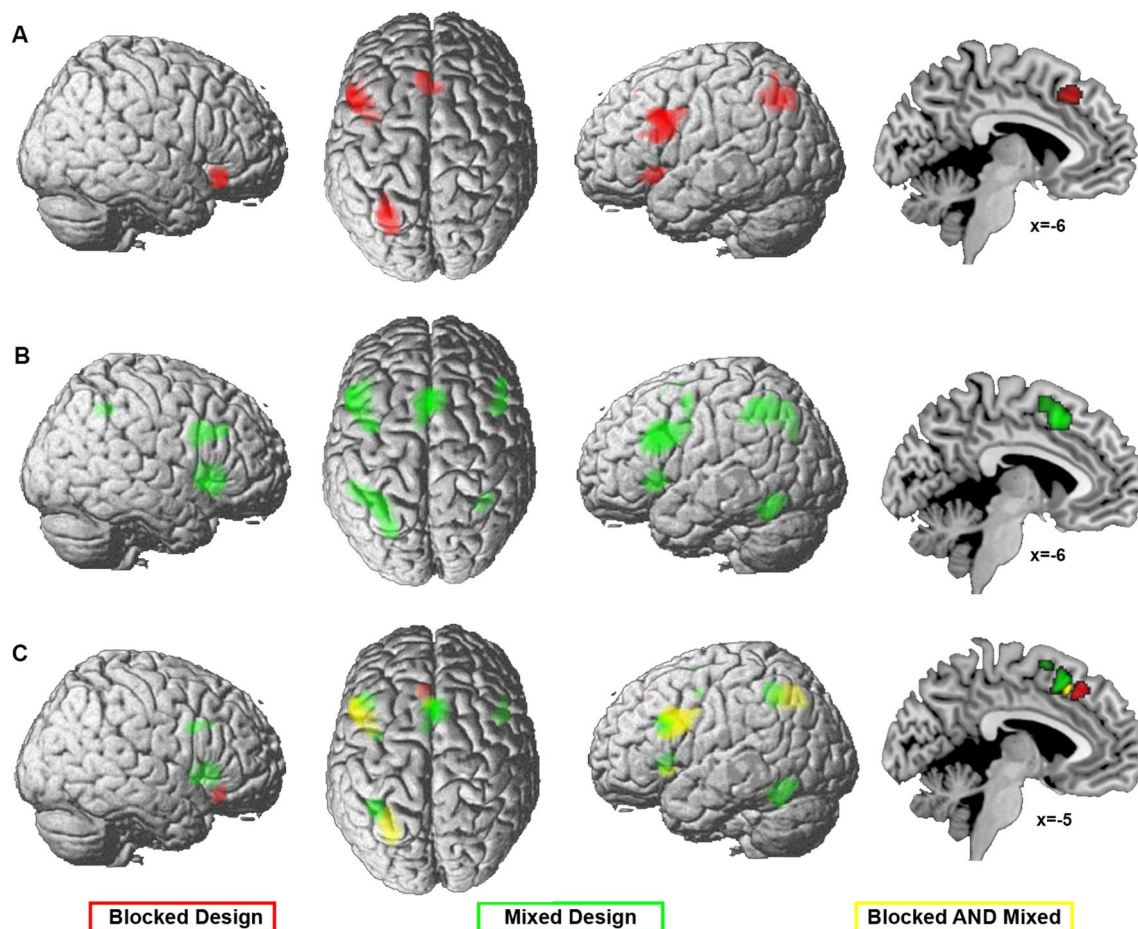


Fig. 3 Meta-analyses across color-word Stroop experiments using **A** blocked (26 experiments) or **B** mixed modes (36 experiments) of presenting incongruent and congruent/neutral stimuli. **C** The results of the conjunction (yellow) and contrast (red/green: stronger convergence for blocked/mixed designs respectively) between the meta-

analyses across experiments of either design. The color intensity of the renderings (the first three columns) reflects the distance from the brain surface (lower intensity denoting further distance from the surface)

was found in the left inferior/middle frontal gyrus and IFJ, left mIPS (hIP3), left dPMC, and left aINS/orbitofrontal cortex (Fig. 7B in green), while other types of Stroop showed stronger convergence in left aIPS/IPL and right aINS (Fig. 7B in blue).

Emotional and other types of Stroop tasks did not reveal any additional overlap beyond that found in the three-way conjunction. Emotional Stroop tasks more consistently recruited left mIPS (hIP3) and left ventral dPMC (Fig. 7C, in red), whereas no stronger convergence was found for other types of Stroop tasks.

In summary, all three types of stimulus material elicited consistent interference-related activity in regions of the salience network, in particular the pmFC and the right aINS. Emotional Stroop was found to be specifically associated with ventral dPMC and color-word Stroop with the left middle frontal cortex, while other types of Stroop showed

generally less convergence as well as stronger convergence in aIPS only in comparison to the color-word version.

Follow-up Analysis: Meta-analyses of the Behavioral Stroop Effect

Table 2, part A, provides an overview of all analyses and the results. For the analyses of the behavioral Stroop effect, we first estimated a mean effect size across all experiments. Then, two meta-regression models were calculated, one testing the effects of control condition, presentation design, and additional cognitive demand for color-word Stroop only and one testing the moderator stimulus material for experiments using a congruent condition as control and without additional cognitive demand. At last, sample biases and outlier analyses were performed and all analyses were repeated without outliers.

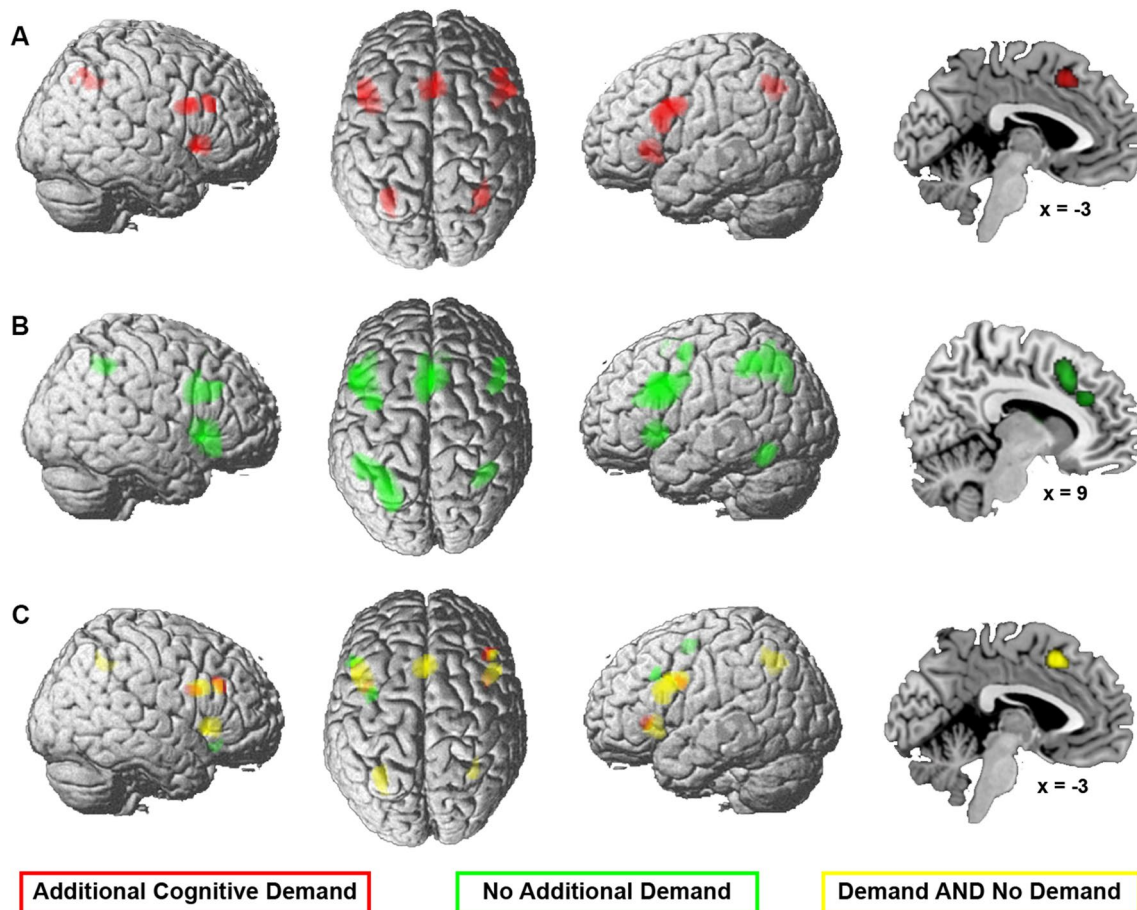


Fig. 4 Meta-analyses across color-word Stroop experiments **A** with an additional cognitive demand component (20 experiments) and **B** without additional demand (66 experiments). **C** The conjunction (yellow) and contrast (red, additional cognitive demand; green, no additional cognitive demand) between both analyses. The color intensity of the renderings (the first three columns) reflects the distance from the brain surface (lower intensity denoting further distance from the surface)

Intercept-Only Model

To estimate the aggregated effect size (ES) estimate across all experiments (77 studies with 68 independent studyIDs and 164 ES), an intercept-only model was calculated, yielding $\bar{g}=0.64$ ($SE=0.04$; 95% confidence interval (CI)=0.56–0.73; degrees of freedom (dfs)=65.7, $p<0.0001$).

Meta-regression Across Color-Word Stroop Experiments

The meta-regression model testing the moderators control condition, presentation design, and additional demand indicated that none of the moderators had a significant effect (39 studyIDs with 112 ES; Table 3; see also Fig. 8 for mean effect sizes for each condition).

Additional analyses that additionally model the moderators type of neutral control condition, interaction between design and control condition, response modality, and

number of different colors can be found in the supplement (tables S5 and S6). These analyses did not reveal any significant effects.

number of different colors can be found in the supplement (tables S5 and S6). These analyses did not reveal any significant effects.

Meta-regression Testing for Effects of the Stimulus Material

The meta-regression (53 studyIDs with 97 ES) testing the moderator “stimulus material” revealed a significant effect (Wald test for testing the overall effect: $F_{2, 31.7}=13.5$, $p<0.001$; Fig. 9 for mean effect sizes of the levels of the moderator). Emotional Stroop tasks exhibited a significantly smaller mean effect size than color-word versions ($\bar{g}_{EMO}=0.41$, $\bar{g}_{CW}=0.78$; $t_{26.4}=-5.1$, $p<0.001$) but not other types of Stroop tasks ($\bar{g}_{Other}=0.63$, $t_{26.4}=1.92$, $p=0.065$). A model using the color-word version as a baseline revealed that color-word and other types of Stroop tasks did not differ significantly ($t_{32.3}=1.16$, $p=0.26$).

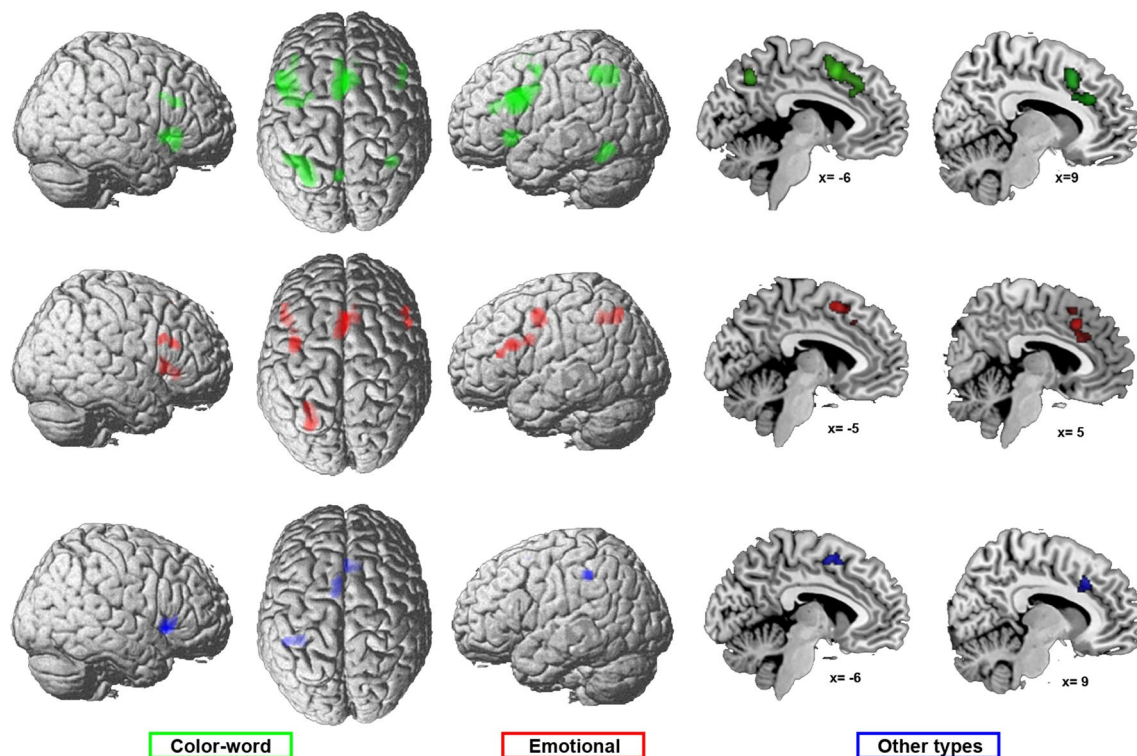


Fig. 5 Main effects of the different types of stimulus material: meta-analyses across experiments using color-word Stroop (green, 42 experiments), emotional (red, 17 experiments), or other types of

Stroop tasks (blue, 20 experiments). The color intensity of the renderings (the first three columns) reflects the distance from the brain surface (lower intensity denoting further distance from the surface)

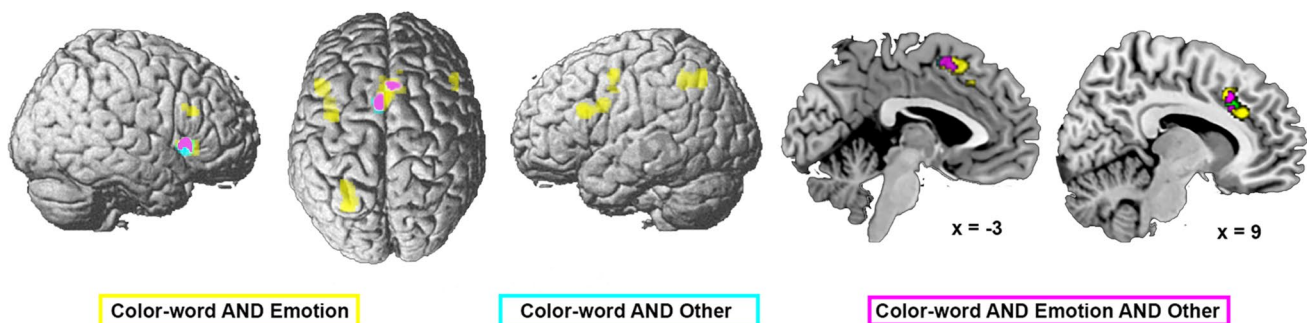


Fig. 6 Conjunctions across the three meta-analyses of the different stimulus materials used, revealing overlap between all three types of material in the posterior medial frontal cortex (pre-SMA and aMCC)

and right aINS (shown in pink). The color intensity of the renderings (the first three columns) reflects the distance from the brain surface (lower intensity denoting further distance from the surface)

Sample Bias

Testing for sampling bias by calculating a meta-regression including SE as a moderator revealed a significant effect for SE (beta=4.50; SE=0.86, $t_{19,1}=5.21$, $p<0.001$), indicating bias in the data. Outlier analyses revealed influential cases for both models (color-word Stroop model: effect size from Kim et al., 2014, and one effect size from Bench et al.,

1993; stimulus material model: two effect sizes from Matthews et al., 2004). Removing those cases did not change any effects of the moderators for color-word Stroop only. When removing outliers from the meta-regression testing for the type of stimulus material, not only emotional Stroop tasks ($t_{26,4} = -5.2$, $p<0.001$) but also other types ($t_{30} = -2.2$, $p=0.03$) differed significantly from color-word stimulus material.

Fig. 7 Differences in convergence between the three stimulus material-specific meta-analyses. Contrast analyses between **A** color-word and emotional task versions, **B** color-word and other Stroop versions and **C** emotional and other Stroop versions. Stronger convergence for color-word versions is shown in green, for emotional versions in red, and for other Stroop versions in blue. The color intensity of the renderings (the first three columns) reflects the distance from the brain surface (lower intensity denoting further distance from the surface)

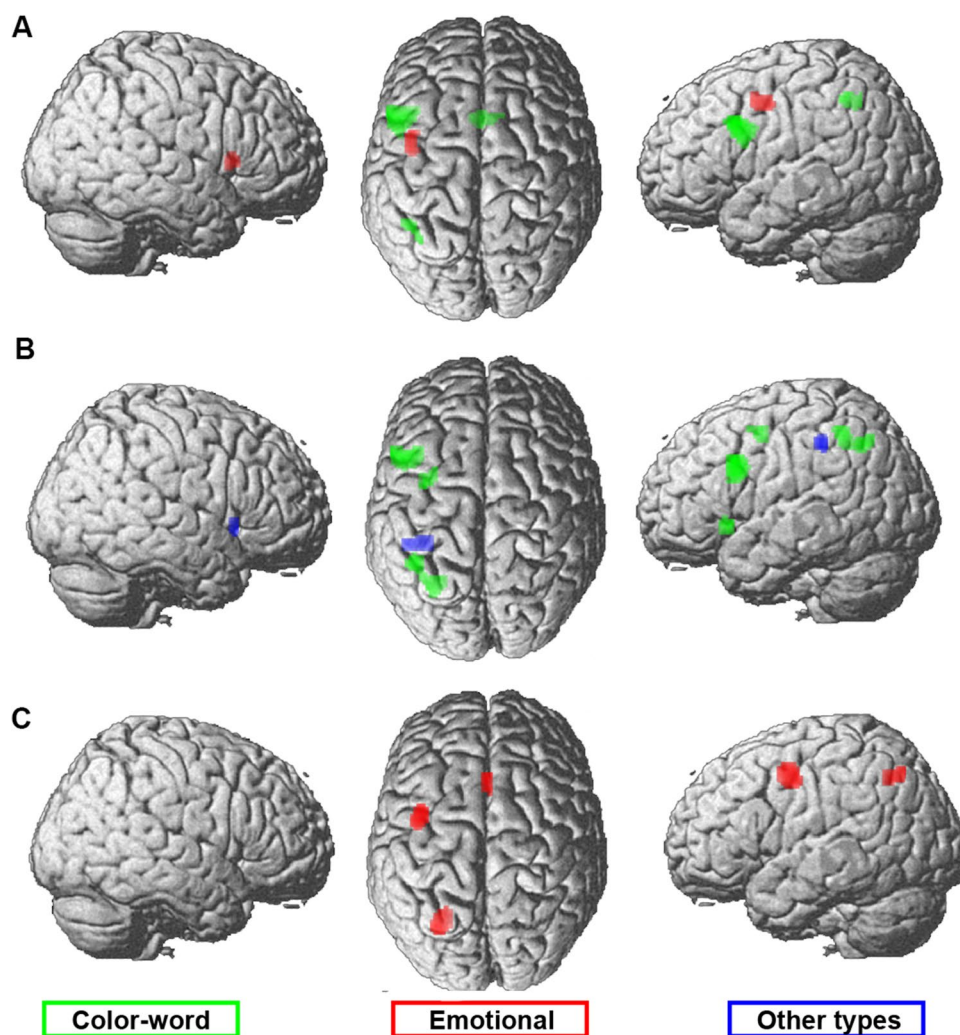


Table 3 Results of the meta-regression across color-word Stroop tasks revealing no significant effect of the moderators control condition, presentation design, and additional cognitive demand

	Estimate	SE	<i>t</i>	df	<i>p</i>	95% CI
Intercept	0.81	0.12	6.7	15.5	<0.0001	0.56–1.07
Control condition	−0.02	0.09	−0.25	28.6	0.805	−0.20–0.15
Design	−0.09	0.13	−0.74	28.6	0.465	−0.35–0.17
Demand	−0.09	0.15	0.57	12.9	0.581	−0.42–0.24

Summarizing the results of the meta-analyses across behavioral Stroop effects, none of the factors investigated (control condition, design, additional demand) had an effect on the size of the incongruence effect during the color-word Stroop task. However, the type of stimulus material significantly modulated the size of the effect, with emotional and other Stroop-type tasks (after removing outliers) leading to smaller interference than observed for the color-word Stroop task version.

Discussion

We performed meta-analyses to investigate the modulatory effects of different task variations on Stroop-type interference as reflected in performance and brain activation. When we investigated the classic color-word Stroop task only, we observed no significant behavioral effects of any variations. However, when comparing to other Stroop-like tasks, we

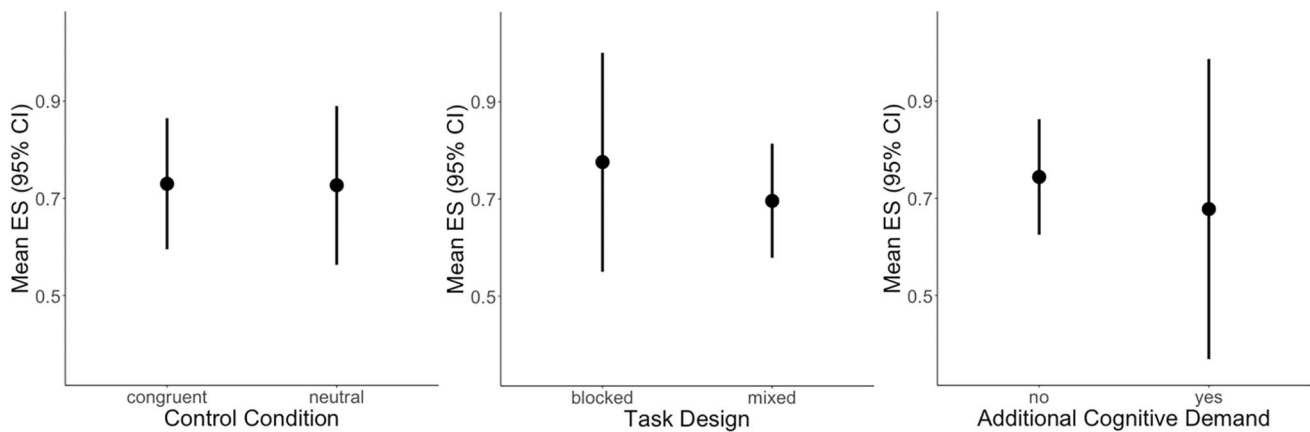


Fig. 8 Illustration of the effect sizes of the color-word Stroop task for each (level of) the moderators control condition, presentation design, and additional cognitive demand. No significant effects were found for any of the moderators

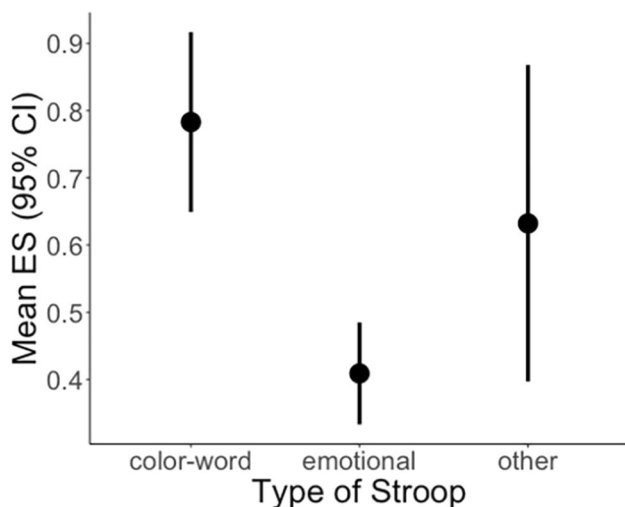


Fig. 9 Illustration of the effect sizes of Stroop interference for the different types of Stroop stimulus material, with a smaller effect for emotional than color-word material

found significant modulations of the behavioral Stroop effect by stimulus material, with the strongest behavioral interference effect observed for the classic color-word Stroop version. On the neural level, left-sided regions of the multiple-demand (MD) network were consistently recruited across variations of the color-word Stroop tasks, with some differences in convergence for right-sided MD regions as well as the fusiform gyrus and orbitofrontal cortex. In addition, when looking at stimulus material other than that used in the color-word and emotional Stroop variants, the neural correlates of the interference effect differed. These results lead to different conclusions about the behavioral and neural interference effect with different neural mechanisms being associated with variations in experimental setup which, however, lead to similar behavior. The current

meta-analytic study thus provides crucial insights for clinical and cognitive neuroscience as differences in neural mechanisms may affect the sensitivity of a specific task version in detecting a particular cognitive (dis)ability and highlights the notion that it is quite difficult to generalize effects beyond the specific task version used.

The Behavioral Stroop Interference Effect

Aggregation of effect sizes of the behavioral interference effect revealed a medium effect size across all conditions for Stroop tasks conducted in neuroimaging environments. Importantly, this effect was quite consistent (i.e., similarly strong) across different variations of color-word Stroop tasks, as effect sizes were independent of the control conditions, presentation design, and the presence of additional cognitive demand. These results side with the idea that the exclusive driver of the behavioral Stroop effect is the higher-order conflict that is common to all Stroop-type tasks, which arises when two semantically related stimulus dimensions are incongruent to each other. The influence of variations of the classic color-word version on Stroop-type interference has been rarely studied systematically, and conclusions could often only be drawn indirectly. For example, we assumed that the type of control condition (congruent vs. neutral) would affect the interference effect size, as benefits for reaction time (RT) from facilitation by congruent (but not neutral) stimuli seem intuitive and have been previously reported. However, it has also been argued that the facilitation effect is much smaller than the interference effect (Macleod, 1991). Our results now add to this discussion by indicating that the presumably small effects of facilitation per se do not consistently lead to a significantly stronger Stroop interference effect when incongruent conditions are compared to congruent rather than neutral conditions.

For blocked versus mixed designs, previous studies reported inconsistent results, with some pointing to stronger interference effects for mixed (Floden et al., 2011) and others for blocked presentations (Hasshim & Parris, 2018; Salo et al., 2001). Our results, summarizing effects found across many studies, now indicate that none of the previously reported effects reported for design and control condition is generalizable and that the kind of control condition, presentation design, and cognitive demand has no impact on the size of the Stroop effect for color-word tasks in the neuroimaging settings.

The only variation that did make a difference was the type of stimulus material used, with the interference effects of emotional face Stroop tasks being smaller than that of color-word ones. Additionally, when removing outliers, the analysis also revealed a difference between other types and color-word Stroop, with smaller effects observed for the former. This contrasts with Chechko et al. (2013), who reported stronger effects for emotional than non-emotional face-word versions of the task. We assume that the effect reported in Chechko et al. (2013) cannot easily be generalized to other task types but is rather specific to the comparison of emotional to non-emotional face-word Stroop task versions. Unfortunately, we only had a very limited amount of non-emotional face Stroop experiments in our sample and thus included non-emotional face experiments in the mixed category “other types,” precluding separate analysis. In general, the results of the few previous studies that compare effects between tasks are quite inconsistent, with some reporting differences (Banich et al., 2000; Capizzi et al., 2017; Chechko et al., 2013; Hilbert et al., 2014), but others not (Mitchell, 2005; Zoccatelli et al., 2010). Importantly, those previous studies differ in the specific tasks that have been compared as well as the settings (neuroimaging vs. outside of the scanner setting). By aggregating the pertinent neuroimaging literature, our findings further our understanding of which task variations do or do not influence the Stroop interference effect when performed inside the scanner, indicating that different variations of the color-word Stroop task lead to comparable behavioral effects, while the emotional and other Stroop-type tasks should be treated differently.

The Multiple-Demand (MD) Network During Stroop Interference Processing

The MD network, consisting of the lateral frontal cortex, anterior insula (aINS)/frontal operculum, and intraparietal sulcus (IPS) as well as pre-supplementary motor area (pre-SMA) and dorsal anterior cingulate cortex (hereafter jointly labeled pmFC), has been found to be associated with various cognitive challenging tasks. It is therefore assumed to play a major role in establishing general mental control

programs across a wide range of tasks with demand for top-down control (Duncan, 2010, 2013). Accordingly, we would have assumed that the higher-order conflict that all Stroop-type tasks have in common, i.e., the overlap of the two dimensions of the stimulus (Kornblum & Stevens, 2002), is reflected by a consistent involvement of this system. In line with this reasoning and with previous meta-analytic findings on interference processing (Chen et al., 2018a; Cieslik et al., 2015; Huang et al., 2020), the present study found convergence in regions of this brain system across various experiments of the large family of color-word and emotional picture-word Stroop tasks. However, while some of these regions were quite consistently found in (almost) all analyses, there were other regions of the MD network that were modulated by control condition and especially by presentation design and stimulus material. Thus, our study provides a more fine-grained differentiation of the MD system, with some regions being more tightly associated with higher-order conflict that all included experiments have in common (resulting from the overlap of two stimulus dimensions) and others being more related to conflict specific to some task variations.

Influence of the Control Condition

The regions revealed by the two meta-analyses across I > C and I > N contrasts, respectively, are rather similar, which is confirmed by the conjunction analysis showing convergence in bilateral aINS and IPFC, left IPS, and pmFC. The only difference between the two resulting networks was that most of the MD regions of convergence for I > N were, in general, larger.

The more extended and stronger convergence of MD regions for experiments that contrast interference processing against a neutral condition might point to task conflict (MacLeod & MacDonald, 2000) in congruent conditions that are absent/reduced from neutral ones, effectively reducing the difference between the congruent and the target conflict condition (i.e., the incongruent one). Task conflict refers to the simultaneous activation of two or more different task sets, i.e., besides preparation for the task set of color identification, also, the task set of word reading is activated. This concurrent preparation of the two different task sets occurs not only in incongruent conditions but also in congruent ones (Littman et al., 2019; MacLeod & MacDonald, 2000; Parris, 2014). In contrast, neutral conditions using non-words activate the task set of word reading to a lesser extent than congruent and incongruent words (Keha & Kalanthroff, 2023; Parris, 2014). Therefore, task conflict in congruent conditions might potentially also recruit MD regions to some extent, resulting in smaller differences in recruitment of those regions between incongruent and congruent conditions. However, Parris et al. (2019) did not find any evidence

for regions associated with task conflict. Additionally, while others have linked task conflict in color-word Stroop versions especially to the pmFC region (in particular cingulate parts; MacLeod & MacDonald, 2000), our results indicate that rather the whole MD network might be involved. Importantly, when separating the experiments of $I > N$ into those that used non-words as neutral conditions (letters or symbols—therefore, little task conflict) and those that used neutral words (including task conflict), results were similar to when combining all neutral control conditions. Only the left aINS was selectively found in the analysis across neutral control conditions using symbols/letters (vs. neutral words). Thus, task conflict can potentially explain stronger convergence in the left aINS when using a neutral control condition compared to a congruent one but cannot fully explain the broader recruitment of other regions of the MD system when controlling against a neutral condition.

In summary, our results provide some evidence for potentially different mechanisms involved in processing congruent versus neutral conditions, leading to a generally stronger difference in the recruitment of MD regions between incongruent and neutral conditions, as compared to contrasts between incongruent and congruent conditions, which cannot, however, be fully attributed to task conflict.

Influence of Presentation Design

Looking at the results of the two meta-analyses of experiments presenting conditions in blocked or mixed fashion, respectively, reveals that the Stroop effect is associated with convergence in more brain regions when mixed compared to blocked designs are used. Both analyses showed convergence in left IPFC, IPS, aINS, and pmFC, but experiments with mixed (vs. blocked) presentation showed additional convergence in right IPFC, right aINS, and left FFG as well as a generally stronger convergence of all MD regions (with the exception of more anterior pmFC showing stronger convergence in blocked designs). This result is in line with a previous study that found stronger effects for mixing (vs. blocking) conditions in the MD network in Stroop (Floden et al., 2011) as well as flanker tasks (Marini et al., 2016) and more left-sided involvement in a Stroop task with blocked presentation of the Stroop (albeit without not direct comparison to a mixed version, Leung et al., 2000).

The more bilateral involvement for mixed designs might be explained by difficulty and mental load, with increasing recruitment of the MD system as well as involvement of regions of the less dominant (i.e., right for verbal tasks like color-word Stroop) hemisphere as load increases (Shashidhara et al., 2019). Thus, right-lateralized regions might be co-recruited in more challenging conditions, as it is when conditions are mixed, to overcome increased interference.

Interestingly, this seems to lead to similar behavioral interference effects in blocked and mixed designs, despite greater difficulty in the latter. It might be that the recruitment of right-sided regions compensates for mixing costs.

In general, blocked and mixed designs differ in the amount and type of carry-over effects, with blocked designs exhibiting stronger adaptation effects and proactive control (Hasshim & Parris, 2018). For solving color-word Stroop trials in a blocked design, interference can be anticipated and appropriate control settings can be implemented in advance and maintained across the whole block, which is beneficial. Additionally, Kalanthroff et al. (2018) suggested that task conflict is particularly apparent when proactive control (maintenance of attention/control) is low. Stronger (and more bilateral) convergence in the MD system in mixed designs (see Fig. 5) in the present study might therefore be due to stronger task conflict in mixed designs (as proactive control is lower compared to blocked presentation). However, as discussed before, our results regarding the type of control condition do not support the effects of task conflict on most regions of the MD system, except for the left aINS. Interestingly, left aINS was also found to show stronger convergence in mixed than in blocked designs. Thus, together with the results of the control condition, the findings of blocked versus mixed design point to the left aINS in playing a major role in Stroop-type task conflicts.

Another explanation for the differences between blocked and mixed designs might be the modeling of the hemodynamic response. While for mixed presentations, analysis is event-related where every experimental event is convolved with the hemodynamic response model, the analysis of blocked conditions convolves the whole boxcar time course. It might be the case that some regions show an early activation and a late deactivation, which could lead to a cancellation of the response in blocked designs, while event-related analysis might not capture the late response (i.e., the deactivation) and thus activation is predominantly found (Meltzer et al., 2008). However, why this should be especially the case for right-sided regions remains an open question for now.

The pattern of left-sided vs. more bilateral involvement in blocked versus mixed designs, respectively, might explain divergent findings in clinical studies, in which right-sided aberrations are more likely to be found when using a mixed compared to a blocked design. Thus, blocked designs might be particularly sensitive for left-sided impairments, while right-hemispheric dysfunctions might only become apparent when conditions are presented in a mixed format. This is in line with the findings of human lesion studies pointing to especially lesions of IPFC on the left side being associated with Stroop performance decline (Cipolotti et al., 2016; Glascher et al., 2012; Perret, 1974) and the fact that clinical studies primarily use original (clinical) forms of the Stroop

task in which word stimuli are presented in a blocked, list-wise format.

In addition to the generally stronger and more bilateral convergence in the MD for mixed (vs. blocked) presentation designs, we found that the left FFG was more consistently reported when conditions were mixed. The cluster in the FFG corresponds to the fusiform word area (Cohen & Dehaene, 2004; Cohen et al., 2000, 2002; Lorenz et al., 2017), a region involved in word reading (Cohen & Dehaene, 2004; Cohen et al., 2000, 2002). During the performance of the Stroop task, one would expect neural suppression of this region to inhibit word reading (Polk et al., 2008). The fact that we did not find convergence for blocked designs could indicate that suppression worked and word reading was successfully inhibited when the same conditions were presented in a row. However, in mixed designs, suppression might be more difficult, especially in conditions in which word meaning and ink color do not match.

Additionally, we found stronger convergence for blocked designs in the right orbitofrontal cortex, often referred to as part of the ventrolateral prefrontal cortex (vlPFC). Importantly, a study by Egner (2011) suggested that the right vlPFC plays a key role in conflict adaptation by showing that the strength of the behavioral conflict adaptation effects increases with more activation in this region. However, the vlPFC was only found when using an interindividual differences approach, while analyses across groups point to the dorsolateral prefrontal cortex (dlPFC) (Egner & Hirsch, 2005; Egner et al., 2008). The vlPFC might therefore be the main source of regulation processes, while (right) dlPFC is only recruited additionally if adaptation does not work or is difficult (i.e., in poor performers; Egner, 2011). This assumption could explain why vlPFC, but not right dlPFC, is found for blocked designs: in a blocked presentation mode, behavioral adaptation is easier, and after a few trials, even participants with difficulties in adapting to the conflict may manage to adjust, and therefore, no further recruitment of right dlPFC is needed. The results of the current meta-analysis might therefore reflect carry-over effects and, in particular, adaptation induced by blocking conditions, and they additionally indicate that blocked designs may be more sensitive to adaptation effects that are only apparent in mixed designs when using individual-differences approaches.

Furthermore, contrast analyses revealed a differentiation within the pmFC, with stronger convergence in pre-SMA for mixed presentation designs, whereas, for blocked designs, more consistent activation was found in a more anterior cluster. pmFC convergence found in the conjunction was located between those two clusters. This differentiation in pmFC is quite interesting, as it has been previously found that proactive (anticipation and prevention of interference before it occurs) and reactive (transient detection of interference after it occurred) controls were associated with different parts

within the pmFC, with a more anterior region linked to proactive and a more posterior one to reactive control (Burgess & Braver, 2010). Assuming that proactive and reactive controls differ in blocked and mixed designs, with stronger proactive control in blocked and reactive one in mixed designs, this differentiation in the pmFC might therefore reflect this distinction between these two control mechanisms.

In summary, our results show that context alters interference-related processing in the Stroop task, as reflected in the differential recruitment of the MD network during blocked versus mixed presentation modes, potentially leading to a similar behavioral Stroop effect for both designs. Stronger and bilateral recruitment of the MD network in mixed designs might reflect increased cognitive load, while right vlPFC convergence for blocked designs could be the result of adaptation. Additionally, together with the results on the effects of the type of control condition, our findings point to the left aINS as playing a role in task conflict. Finally, a differentiation within pmFC between the two task designs most likely reflects differences in the exertion of pro- and reactive control.

Influence of Stimulus Material

This meta-analytic study showed that the stimulus material used to induce Stroop-type conflict (i.e., color-word, emotional, or other Stroop-like tasks) has some impact on which interference-related regions are consistently found across experiments. The only overlap between all three categories of Stroop tasks employed in this study was found in the pmFC (pre-SMA and anterior midcingulate cortex) and right aINS. Both regions are part of the MDN but have in particular been described as forming the salience network (Seeley et al., 2007), playing a major role in detecting salient stimuli and implementing and maintaining the appropriate task set via initiating switches between task-relevant and task-irrelevant networks (Dosenbach et al., 2006; Menon & Uddin, 2010; Sridharan et al., 2008). Consistent involvement of the salience network across all three material-specific categories of the Stroop tasks therefore indicates that the detection of (salient) stimulus conflict and the (re)activation of the instructed task set to overcome inadequate response tendencies are implemented in a similar, material-independent way.

Interestingly, our results point to some distinctions within left IPS with regard to stimulus material. In particular, color-word and emotional Stroop experiments primarily converged in caudal parts of the anterior IPS (in particular hIP1, hIP3, and hIP6; Choi et al., 2006; Scheperjans et al., 2008a, 2008b), which we name “middle IPS” (mIPS). In contrast, for other types of Stroop tasks, convergence was mainly and more consistently found in the most anterior part of IPS (hIP2; Choi et al., 2006), as compared to color-word Stroop versions. Studies that investigate the regions associated with

numerical processing often reported regions that overlap with the cluster found in our meta-analyses for other types of Stroop tasks. Zago et al. (2008) reported a stronger role of anterior IPS in working memory tasks with numerical stimuli, as compared to verbal syllables, and Vogel et al. (2017) also pointed to anterior IPS as involved in symbolic number processing across modalities. In general, IPS is held to play a strong role in attentional shifting and stimulus–response mapping (Cieslik et al., 2015; Worringer et al., 2019). During the Stroop task, IPS involvement therefore might reflect the allocation of attention to, and/or the selection of specific stimulus characteristics (Cieslik et al., 2015) like color, location, or numerical quantity of a stimulus. Our results now indicate that within IPS, this attentional shifting might be implemented in a material-specific way. Importantly, when looking at the experiments that contribute to the aIPS cluster in the meta-analysis across other types of stimulus material, all but one experiment used a numerical or counting Stroop task version. Thus, the differentiation within left IPS may point to a shift of recruitment from middle IPS for allocation of attention to color and emotional expression of the stimulus, to more anterior IPS for attentional shifts to magnitude and quantitative properties (like numerical magnitude or font size).

When compared to the other two types of stimulus material, stronger convergence was found in the ventral dorsal premotor cortex (dPMC) for emotional picture-word Stroop tasks and in IPFC for color-word Stroop. In terms of the Stroop task, it has been shown that left dPMC is specifically associated with the regulation of perceptual conflict, while the lateral frontal cortex is more involved in conflict detection and response conflict (Kim et al., 2012). Given the stronger consistency of dPMC across experiments in emotional interference, our results might therefore indicate that resolving perceptual conflict is more relevant in emotional than color-word or other Stroop variants. In turn, stronger convergence for color-word Stroop versions in IPFC indicates that for this task variant, resolution of response conflict may play a stronger role than for the other two categories of stimulus material.

Additionally, it has been suggested that the lateral prefrontal cortex is especially involved when there are differences in the automaticity of the two competing dimensions (Banich, 2019). Importantly, for the color-word Stroop tasks, interference effects only arise when the word meaning is the task-irrelevant dimension, but not when the color of the ink must be ignored (Stroop, 1935). Therefore, the automaticity of the two competing dimensions is unbalanced, leading to a stronger recruitment of the lateral prefrontal cortex. In contrast, for some other Stroop versions but also emotional Stroop, this difference in automaticity of the two competing dimensions is less pronounced. In the numerical Stroop version, for example, behavioral interference effects

are observed when either the physical size or the numerical magnitude is the task-irrelevant dimension (Huang et al., 2012; Tang et al., 2006). Similarly, with emotional Stroop tasks, interference effects are observed for both the identification of facial expression and word reading (Bayer et al., 2018). Thus, the stronger convergence observed in IPFC for color-word Stroop tasks might be due to the fact that there is more competition between the automaticity of the two competing dimensions and, therefore, more need for top-down control.

In summary, our meta-analytic results regarding the impact of different types of stimulus material on the neural correlates of Stroop-like interference processing highlight the salience network as a core system for dealing with such conflicts. Additionally, recruitment of the MD network is to some extent material-specific, suggesting that partially different control mechanisms are recruited for different Stroop variants.

Limitations and Outlook

It must be noted that the behavioral effects reported here can only be generalized to settings typical of neuroimaging experiments. Previous work points to differences in reaction time for studies performed in and outside the scanner (Koch et al., 2003; Koten et al., 2013; van Maanen et al., 2016). Therefore, outside the scanner environment, these effects might look different, and a next step would be to test the influence of task variations on the behavioral Stroop effect in standard laboratory and clinical routine settings.

Furthermore, it should be mentioned that the “other” category of stimulus material comprised more than one Stroop variant. This heterogeneity was due to the limited number of available fMRI experiments for each version, which did not allow for forming more material-specific categories. Therefore, some effects (or their absence) observed for this mixed category of other types of Stroop tasks might be driven by only a specific subgroup of tasks. For the brain regions where convergence was found, our analysis of study contributions indicated that there was not a specific type of task driving the results. However, for regions where no convergence was found in the category of other types of Stroop (e.g., dIPFC), it is unclear if this is potentially due to heterogeneity or one specific type of Stroop. Nevertheless, if the Stroop effect was independent of stimulus material, we would still expect convergence in the same regions as observed for color-word and emotional Stroop versions. Therefore, future studies (single fMRI studies directly comparing different task variants or meta-analyses performed once more experiments have become available) should specifically compare different task types now subsumed in this mixed group.

Furthermore, the Eggers regression test indicated that there is a sampling bias in the reported effect sizes of the behavioral Stroop effect. However, this is not surprising as we pre-selected the ES by focusing on published neuroimaging studies. Additionally, our focus was not on determining the overall effect size. Rather, we were mainly interested in examining whether and how different moderators influence the size of the Stroop interference effect in task versions implemented in neuroimaging settings. We therefore already assessed factors that might explain the heterogeneity of effects and sampling bias.

Additionally, it has to be acknowledged that brain signals can be confounded by reaction time variations (Mumford et al., 2023) and that presumably conflict-related activations can be explained by the time spent on the task. Thus, some neural effects of task variations found in our meta-analysis could potentially not necessarily reflect differences in conflict processing per se but rather be an effect of reaction time.

Besides the stimulus material, presentation design, control condition, and additional cognitive demand, there also are other factors (like the probability of different conditions, stimulus set size, response modality) that might influence the size of the Stroop effect as well as its neural correlates (for review, see Macleod, 1991). However, we only investigated task and experimental variations that were used in a sufficient number of studies.

Conclusion

Overall, our results suggest that different neural mechanisms are associated with variations in the experimental setup of Stroop-like tasks, which, however, lead to similar interference effects on the behavioral level. In line with the view of a “many-to-one mapping” (Westlin et al., 2023), this suggests that the seemingly unitary behavioral costs of Stroop-type conflicts may arise from partly different processing mechanisms, depending on contextual factors. This is especially true for differences in presentation design, where mixed presentations (as compared to blocked designs) recruit the MD network more strongly and more bilaterally, as well as for variations in stimulus material, which differ in the recruitment of parietal, lateral frontal, and dorsal premotor cortex. This therefore highlights that Stroop-type neuroimaging experiments as well as applications of Stroop-like tasks in clinical, occupational, or other diagnostic settings should be carefully planned and interpreted, as task variations influence the set of brain regions recruited. Although behavioral interference effects were, at the level of group averages, hardly affected by task variations, the differences in neural mechanisms may make a given task version more or less sensitive to particular cognitive ability differences. Information on differences in the brain regions recruited for different task versions is thus of high relevance

for clinical neuropsychology, as the version used should recruit the brain region of interest (e.g., mixed presentation mode is recommended when not only left-sided dysfunctions are to be detected). In cognitive neuroscience, in turn, the divergence of behavioral and neuroimaging effects might render it particularly difficult to find brain-behavior relationships that are generalizable across different task versions. Ultimately, our results question the meaningfulness of using “Stroop task” in neuroimaging research and applied settings as an umbrella term for such a wide variety of flavors.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11065-024-09647-1>.

Acknowledgements We would like to thank all contacted authors who replied to our emails and who contributed additional information and results not explicitly reported in the original publication.

Author Contribution VM: conceptualization and design, data acquisition, analysis, writing and revision of the manuscript; EC: conceptualization, analysis, writing and revision of the manuscript; LF: data acquisition, writing and revision of the manuscript; ST: data acquisition, analysis, writing and revision of the manuscript; AS: methodology, writing and revision of the manuscript; TA: data acquisition; writing and revision of the manuscript; CF: data acquisition, writing and revision of the manuscript; SE: methodology, writing and revision of the manuscript; RL: conceptualization and design, analysis, writing and revision of the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, EI 816/11–1), the National Institute of Mental Health (R01-MH074457), and the Helmholtz Portfolio Theme “Supercomputing and Modeling for the Human Brain.”

Data Availability Analysis code of neuroimaging and behavioral meta-analyses can be found on the open science framework (OSF; https://osf.io/dt3kj/?view_only=Scheperjans97bb53574583b1a0dda978f7f341); result files of the ALE meta-analysis are available at ANIMA (<https://anima.fz-juelich.de/>).

Declarations

Conflict of Interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

References marked with asterisk indicate studies included in the meta-analysis: **included in neuroimaging and behavioral meta-analysis; *included in neuroimaging meta-analysis only

- *Adleman, N. E., Menon, V., Blasey, C. M., White, C. D., Warsofsky, I. S., Glover, G. H., & Reiss, A. L. (2002). A developmental fMRI study of the Stroop color-word task. *Neuroimage*, 16(1), 61–75. <https://doi.org/10.1006/nimg.2001.1046>
- **Agostini, A., Ballotta, D., Righi, S., Moretti, M., Bertani, A., Scarcelli, A., Sartini, A., Ercolani, M., Nichelli, P., Campieri, M., & Benuzzi, F. (2017). Stress and brain functional changes in patients with Crohn's disease: A functional magnetic resonance imaging study. *Neurogastroenterology and Motility*, 29(10), 1–10. <https://doi.org/10.1111/nmo.13108>
- Algom, D., Chajut, E., & Lev, S. (2004). A rational look at the emotional Stroop phenomenon: A generic slowdown, not a Stroop effect. *Journal of Experimental Psychology: General*, 133(3), 323–338. <https://doi.org/10.1037/0096-3445.133.3.323>
- Algom, D., Fitousi, D., & Chajut, E. (2022). Can the Stroop effect serve as the gold standard of conflict monitoring and control? *A conceptual critique. Memory and Cognition*, 50(5), 883–897. <https://doi.org/10.3758/s13421-021-01251-5>
- **Almdahl, I. S., Martinussen, L. J., Agartz, I., Hugdahl, K., & Korsnes, M. S. (2021). Inhibition of emotions in healthy aging: Age-related differences in brain network connectivity. *Brain and Behavior*, 11(5), e02052. <https://doi.org/10.1002/brb3.2052>
- **Ansari, D., Fugelsang, J. A., Dhital, B., & Venkatraman, V. (2006). Dissociating response conflict from numerical magnitude processing in the brain: An event-related fMRI study. *Neuroimage*, 32(2), 799–805. <https://doi.org/10.1016/j.neuroimage.2006.04.184>
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73(1), 3–25. <https://doi.org/10.1037/amp0000191>
- Assem, M., Shashidhara, S., Glasser, M. F., & Duncan, J. (2022). Precise topology of adjacent domain-general and sensory-biased regions in the human brain. *Cerebral Cortex*, 32(12), 2521–2537. <https://doi.org/10.1093/cercor/bhab362>
- Augustinova, M., Parris, B. A., & Ferrand, L. (2019). The loci of Stroop interference and facilitation effects with manual and vocal responses. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.01786>
- **Bang, L., Ro, O., & Endestad, T. (2016). Amygdala alterations during an emotional conflict task in women recovered from anorexia nervosa. *Psychiatry Research Neuroimaging*, 248, 126–133. <https://doi.org/10.1016/j.psychres.2015.12.008>
- Banich, M. T. (2019). The Stroop effect occurs at multiple points along a cascade of control: Evidence from cognitive neuroscience approaches. *Frontiers in Psychology*, 10, 2164. <https://doi.org/10.3389/fpsyg.2019.02164>
- Banich, M. T., Milham, M. P., Atchley, R., Cohen, N. J., Webb, A., Wszalek, T., Kramer, A. F., Liang, Z. P., Wright, A., Shenker, J., & Magin, R. (2000). fMRI studies of Stroop tasks reveal unique roles of anterior and posterior brain systems in attentional selection. *Journal of Cognitive Neuroscience*, 12(6), 988–1000. <https://doi.org/10.1162/08989290051137521>
- *Banich, M. T., Milham, M. P., Jacobson, B. L., Webb, A., Wszalek, T., Cohen, N. J., & Kramer, A. F. (2001). Attentional selection and the processing of task-irrelevant information: Insights from fMRI examinations of the Stroop task. *Progress in Brain Research*, 134, 459–470. [https://doi.org/10.1016/s0079-6123\(01\)34030-x](https://doi.org/10.1016/s0079-6123(01)34030-x)
- *Barkley-Levenson, E., Xue, F., Drouman, V., Miller, L. C., Smith, B. J., Jeong, D., Lu, Z. L., Bechara, A., & Read, S. J. (2018). Prefrontal cortical activity during the Stroop task: New insights into the why and the who of real-world risky sexual behavior. *Annals of Behavioral Medicine*, 52(5), 367–379. <https://doi.org/10.1093/abm/kax019>
- **Basten, U., Stelzel, C., & Fiebach, C. J. (2011). Trait anxiety modulates the neural efficiency of inhibitory control. *Journal of Cognitive Neuroscience*, 23(10), 3132–3145. https://doi.org/10.1162/jocn_a_00003
- **Bayer, M., Rubens, M. T., & Johnstone, T. (2018). Simultaneous EEG-fMRI reveals attention-dependent coupling of early face processing with a distributed cortical network. *Biological Psychology*, 132, 133–142. <https://doi.org/10.1016/j.biopsycho.2017.12.002>
- **Becker, T. M., Kerns, J. G., Macdonald, A. W., 3rd, & Carter, C. S. (2008). Prefrontal dysfunction in first-degree relatives of schizophrenia patients during a Stroop task. *Neuropsychopharmacology*, 33(11), 2619–2625. <https://doi.org/10.1038/sj.npp.1301673>
- **Bench, C. J., Frith, C. D., Grasby, P. M., Friston, K. J., Pauls, E., Frackowiak, R. S., & Dolan, R. J. (1993). Investigations of the functional anatomy of attention using the Stroop test. *Neuropsychologia*, 31(9), 907–922. [https://doi.org/10.1016/0028-3932\(93\)90147-r](https://doi.org/10.1016/0028-3932(93)90147-r)
- Birn, R. M., Cox, R. W., & Bandettini, P. A. (2002). Detection versus estimation in event-related fMRI: Choosing the optimal stimulus timing. *Neuroimage*, 15(1), 252–264. <https://doi.org/10.1006/nimg.2001.0964>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to meta-analysis. Wiley. <https://doi.org/10.1002/9780470743386>
- Boucart, M., Mobarek, N., Cuervo, C., & Danion, J. M. (1999). What is the nature of increased Stroop interference in schizophrenia? *Acta Psychologica*, 101(1), 3–25. [https://doi.org/10.1016/S0001-6918\(98\)00037-7](https://doi.org/10.1016/S0001-6918(98)00037-7)
- Braga, P. L. G., Henrique, J. S., Almeida, S. S., Arida, R. M., & Gomes da Silva, S. (2022). Factors affecting executive function performance of Brazilian elderly in the Stroop test. *Brazilian Journal of Medical and Biological Research*, 55, e11917. <https://doi.org/10.1590/1414-431X2022e11917>
- *Brass, M., Derrfuss, J., & von Cramon, D. Y. (2005). The inhibition of imitative and overlearned responses: A functional double dissociation. *Neuropsychologia*, 43(1), 89–98. <https://doi.org/10.1016/j.neuropsychologia.2004.06.018>
- Braver, T. S., Gray, J. R., & Burgess, G. C. (2008). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In Andrew, C. (Ed.), *Variation in Working Memory*. Oxford Academic. <https://doi.org/10.1093/acprof:oso/9780195168648.003.0004>
- Burgess, G. C., & Braver, T. S. (2010). Neural mechanisms of interference control in working memory: Effects of interference expectancy and fluid intelligence. *PloS One*, 5(9), e12861. <https://doi.org/10.1371/journal.pone.0012861>
- Bush, G., Whalen, P. J., Rosen, B. R., Jenike, M. A., McInerney, S. C., & Rauch, S. L. (1998). The counting Stroop: An interference task specialized for functional neuroimaging—validation study with functional MRI. *Human Brain Mapping*, 6(4), 270–282. [https://doi.org/10.1002/\(SICI\)1097-0193\(1998\)6:4%3c270::AID-HBM6%3e3.0.CO;2-0](https://doi.org/10.1002/(SICI)1097-0193(1998)6:4%3c270::AID-HBM6%3e3.0.CO;2-0)
- Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A. R., Langner, R., & Eickhoff, S. B. (2012). Parsing the neural correlates of moral cognition: ALE meta-analysis on morality,

- theory of mind, and empathy. *Brain Structure & Function*, 217(4), 783–796. <https://doi.org/10.1007/s00429-012-0380-y>
- Capizzi, M., Ambrosini, E., & Vallesi, A. (2017). Individual differences in verbal and spatial Stroop tasks: Interactive role of handedness and domain. *Frontiers in Human Neuroscience*, 11, 545. <https://doi.org/10.3389/fnhum.2017.00545>
- **Carter, C. S., Mintun, M., & Cohen, J. D. (1995). Interference and facilitation effects during selective attention: An H215O PET study of Stroop task performance. *Neuroimage*, 2(4), 264–272. <https://doi.org/10.1006/nimg.1995.1034>
- Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain [Meta-Analysis Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. *Neuroimage*, 50(3), 1148–1167. <https://doi.org/10.1016/j.neuroimage.2009.12.112>
- Chase, H. W., Kumar, P., Eickhoff, S. B., & Dombrovski, A. Y. (2015). Reinforcement learning models and their neural correlates: An activation likelihood estimation meta-analysis. *Cognitive, Affective & Behavioral Neuroscience*. <https://doi.org/10.3758/s13415-015-0338-7>
- **Chechko, N., Kellermann, T., Zvyagintsev, M., Augustin, M., Schneider, F., & Habel, U. (2012). Brain circuitries involved in semantic interference by demands of emotional and non-emotional distractors. *PLoS ONE*, 7(5), e38155. <https://doi.org/10.1371/journal.pone.0038155>
- **Chechko, N., Wehrle, R., Erhardt, A., Holsboer, F., Czisch, M., & Samann, P. G. (2009). Unstable prefrontal response to emotional conflict and activation of lower limbic structures and brainstem in remitted panic disorder. *PloS One*, 4(5), e5537. <https://doi.org/10.1371/journal.pone.0005537>
- **Chechko, N., Augustin, M., Zvyagintsev, M., Schneider, F., Habel, U., & Kellermann, T. (2013). Brain circuitries involved in emotional interference task in major depression disorder. *Journal of Affective Disorders*, 149(1–3), 136–145. <https://doi.org/10.1016/j.jad.2013.01.013>
- Chen, Z., Lei, X., Ding, C., Li, H., & Chen, A. (2013). The neural mechanisms of semantic and response conflicts: An fMRI study of practice-related effects in the Stroop task. *NeuroImage*, 66, 577–584. <https://doi.org/10.1016/j.neuroimage.2012.10.028>
- Chen, T., Becker, B., Camilleri, J., Wang, L., Yu, S., Eickhoff, S. B., & Feng, C. (2018). A domain-general brain network underlying emotional and cognitive interference processing: Evidence from coordinate-based and functional connectivity meta-analyses. *Brain Structure & Function*, 223(8), 3813–3840. <https://doi.org/10.1007/s00429-018-1727-9>
- *Chen, Z., Zhao, X., Fan, J., & Chen, A. (2018). Functional cerebral asymmetry analyses reveal how the control system implements its flexibility. *Human Brain Mapping*, 39(12), 4678–4688. <https://doi.org/10.1002/hbm.24313>
- Choi, H. J., Zilles, K., Mohlberg, H., Schleicher, A., Fink, G. R., Armstrong, E., & Amunts, K. (2006). Cytoarchitectonic identification and probabilistic mapping of two distinct areas within the anterior ventral bank of the human intraparietal sulcus. *Journal of Comparative Neurology*, 495(1), 53–69. <https://doi.org/10.1002/cne.20849>
- Cieslik, E. C., Mueller, V. I., Eickhoff, C. R., Langner, R., & Eickhoff, S. B. (2015). Three key regions for supervisory attentional control: Evidence from neuroimaging meta-analyses. *Neuroscience and Biobehavioral Reviews*, 48, 22–34. <https://doi.org/10.1016/j.neubiorev.2014.11.003>
- Cipolotti, L., Healy, C., Spano, B., Lecce, F., Biondo, F., Robinson, G., Chan, E., Duncan, J., Shallice, T., & Bozzali, M. (2016). Strategy and suppression impairments after right lateral prefrontal and orbito-frontal lesions. *Brain*, 139(Pt 2), e10. <https://doi.org/10.1093/brain/awv269>
- Clark, V. P. (2012). A history of randomized task designs in fMRI. *Neuroimage*, 62(2), 1190–1194. <https://doi.org/10.1016/j.neuroimage.2012.01.010>
- **Coderre, E. L., & van Heuven, W. J. (2013). Modulations of the executive control network by stimulus onset asynchrony in a Stroop task. *BMC Neuroscience*, 14, 79. <https://doi.org/10.1186/1471-2202-14-79>
- *Coderre, E. L., Filippi, C. G., Newhouse, P. A., & Dumas, J. A. (2008). The Stroop effect in kana and kanji scripts in native Japanese speakers: An fMRI study. *Brain and Language*, 107(2), 124–132. <https://doi.org/10.1016/j.bandl.2008.01.011>
- Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: The case for the visual word form area. *Neuroimage*, 22(1), 466–476. <https://doi.org/10.1016/j.neuroimage.2003.12.049>
- Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M. A., & Michel, F. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, 123(Pt 2), 291–307.
- Cohen, L., Lehericy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the visual word form area. *Brain*, 125(Pt 5), 1054–1069. <https://doi.org/10.1093/brain/awf094>
- Dalrymple-Alford, E. C., & Budayer, B. (1966). Examination of some aspects of the Stroop color-word test. *Perceptual and Motor Skills*, 23(3), 1211–1214. <https://doi.org/10.2466/pms.1966.23.3f.1211>
- De Houwer, J., & Hermans, D. (1994). Differences in the affective processing of words and pictures. *Cognition and Emotion*, 8, 1–20. <https://doi.org/10.1080/02699939408408925>
- Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System (D-KEFS)*. Psychological Corporation.
- Demakis, G. J. (2004). Frontal lobe damage and tests of executive processing: A meta-analysis of the category test, Stroop test, and trail-making test. *Journal of Clinical and Experimental Neuropsychology*, 26(3), 441–450. <https://doi.org/10.1080/13803390490510149>
- *DeVito, E. E., Worhunsky, P. D., Carroll, K. M., Rounsaville, B. J., Kober, H., & Potenza, M. N. (2012). A preliminary study of the neural effects of behavioral therapy for substance use disorders. *Drug and Alcohol Dependence*, 122(3), 228–235. <https://doi.org/10.1016/j.drugalcdep.2011.10.002>
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Dimoska-Di Marco, A., McDonald, S., Kelly, M., Tate, R., & Johnstone, S. (2011). A meta-analysis of response inhibition and Stroop interference control deficits in adults with traumatic brain injury (TBI). *Journal of Clinical and Experimental Neuropsychology*, 33(4), 471–485. <https://doi.org/10.1080/13803395.2010.533158>
- Dintica, C. S., Hoang, T., Allen, N., Sidney, S., & Yaffe, K. (2022). The metabolic syndrome is associated with lower cognitive performance and reduced white matter integrity in midlife: The CARDIA study. *Frontiers in Neuroscience*, 16, 942743. <https://doi.org/10.3389/fnins.2022.942743>
- Dosenbach, N. U., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., Burgund, E. D., Grimes, A. L., Schlaggar, B. L., & Petersen, S. E. (2006). A core system for the implementation of task sets. *Neuron*, 50(5), 799–812. <https://doi.org/10.1016/j.neuron.2006.04.031>
- Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, 14(4), 172–179. <https://doi.org/10.1016/j.tics.2010.01.004>

- Duncan, J. (2013). The structure of cognition: Attentional episodes in mind and brain. *Neuron*, 80(1), 35–50. <https://doi.org/10.1016/j.neuron.2013.09.015>
- Egner, T. (2011). Right ventrolateral prefrontal cortex mediates individual differences in conflict-driven cognitive control. *Journal of Cognitive Neuroscience*, 23(12), 3903–3913. https://doi.org/10.1162/jocn_a_00064
- Egner, T., & Hirsch, J. (2005). The neural correlates and functional integration of cognitive control in a Stroop task. *Neuroimage*, 24(2), 539–547. <https://doi.org/10.1016/j.neuroimage.2004.09.007>
- Egner, T., Etkin, A., Gale, S., & Hirsch, J. (2008). Dissociable neural systems resolve conflict from emotional versus nonemotional distracters. *Cerebral Cortex*, 18(6), 1475–1484. <https://doi.org/10.1093/cercor/bhm179>
- Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. *Human Brain Mapping*, 30(9), 2907–2926. <https://doi.org/10.1002/hbm.20718>
- Eickhoff, S. B., Bzdok, D., Laird, A. R., Roski, C., Caspers, S., Zilles, K., & Fox, P. T. (2011). Co-activation patterns distinguish cortical modules, their connectivity and functional differentiation [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. *Neuroimage*, 57(3), 938–949. <https://doi.org/10.1016/j.neuroimage.2011.05.021>
- Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., & Fox, P. T. (2012). Activation likelihood estimation meta-analysis revisited [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. *Neuroimage*, 59(3), 2349–2361. <https://doi.org/10.1016/j.neuroimage.2011.09.017>
- Eickhoff, S. B., Nichols, T. E., Laird, A. R., Hoffstaedter, F., Amunts, K., Fox, P. T., Bzdok, D., & Eickhoff, C. R. (2016). Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. *Neuroimage*. <https://doi.org/10.1016/j.neuroimage.2016.04.072>
- *Fan, J., Flombaum, J. I., McCandliss, B. D., Thomas, K. M., & Posner, M. I. (2003). Cognitive and brain consequences of conflict. *Neuroimage*, 18(1), 42–57. <https://doi.org/10.1006/nimg.2002.1319>
- **Fechir, M., Gamer, M., Blasius, I., Bauermann, T., Breimhorst, M., Schlindwein, P., Schlereth, T., & Birklein, F. (2010). Functional imaging of sympathetic activation during mental stress. *Neuroimage*, 50(2), 847–854. <https://doi.org/10.1016/j.neuroimage.2009.12.004>
- Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 110(41), 16616–16621. <https://doi.org/10.1073/pnas.1315235110>
- Feng, C., Becker, B., Huang, W., Wu, X., Eickhoff, S. B., & Chen, T. (2018). Neural substrates of the emotion-word and emotional counting Stroop tasks in healthy and clinical populations: A meta-analysis of functional brain imaging studies. *NeuroImage*, 173, 258–274. <https://doi.org/10.1016/j.neuroimage.2018.02.023>
- Fisher, Z., & Tipton, E. (2015). *robumeta: An R-package for robust variance estimation in meta-analysis* from <http://arxiv.org/abs/1503.02220>
- Fisher, Z., Tipton, E., & Zhipeng, H. (2017). *robumeta: Robust variance meta-regression (2.0)*. <https://github.com/zackfisher/robumeta>. Accessed Aug 2021
- **Fleury, V., Cousin, E., Czernecki, V., Schmitt, E., Lhomme, E., Poncet, A., Fraix, V., Tropres, I., Pollak, P., Krainik, A., & Krack, P. (2014). Dopaminergic modulation of emotional conflict in parkinson's disease. *Frontiers in Aging Neuroscience*, 6, 164. <https://doi.org/10.3389/fnagi.2014.00164>
- Floden, D., Vallesi, A., & Stuss, D. T. (2011). Task context and frontal lobe activation in the Stroop task. *Journal of Cognitive Neuroscience*, 23(4), 867–879. <https://doi.org/10.1162/jocn.2010.21492>
- Frahm, L., Cieslik, E. C., Hoffstaedter, F., Satterthwaite, T. D., Fox, P. T., Langner, R., & Eickhoff, S. B. (2022). Evaluation of thresholding methods for activation likelihood estimation meta-analysis via large-scale simulations. *Human Brain Mapping*, 43(13), 3987–3997. <https://doi.org/10.1002/hbm.25898>
- *George, M. S., Ketter, T. A., Parekh, P. I., Rosinsky, N., Ring, H., Casey, B. J., Trimble, M. R., Horwitz, B., Herscovitch, P., & Post, R. M. (1994). Regional brain activity when selecting a response despite interference: An H2 (15) O PET study of the Stroop and an emotional Stroop. *Human Brain Mapping*, 1(3), 194–209. <https://doi.org/10.1002/hbm.460010305>
- *Ghavidel, N., Khodagholi, F., Ahmadiani, A., Khosrowabadi, R., Asadi, S., & Shams, J. (2020). Frontocingulate dysfunction is associated with depression and decreased serum PON1 in methamphetamine-dependent patients. *Neuropsychiatric Disease and Treatment*, 16, 489–499. <https://doi.org/10.2147/NDT.S237528>
- *Gianaros, P. J., Sheu, L. K., Matthews, K. A., Jennings, J. R., Manuck, S. B., & Hariri, A. R. (2008). Individual differences in stressor-evoked blood pressure reactivity vary with activation, volume, and functional connectivity of the amygdala. *Journal of Neuroscience*, 28(4), 990–999. <https://doi.org/10.1523/JNEUROSCI.3606-07.2008>
- Glascher, J., Adolphs, R., Damasio, H., Bechara, A., Rudrauf, D., Calamia, M., Paul, L. K., & Tranel, D. (2012). Lesion mapping of cognitive control and value-based decision making in the prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 109(36), 14681–14686. <https://doi.org/10.1073/pnas.1206608109>
- **Godinez, D. A., McRae, K., Andrews-Hanna, J. R., Smolker, H., & Banich, M. T. (2016). Differences in frontal and limbic brain activation in a small sample of monozygotic twin pairs discordant for severe stressful life events. *Neurobiology of Stress*, 5, 26–36. <https://doi.org/10.1016/j.ynstr.2016.10.002>
- Gotlib, I. H., & McCann, C. D. (1984). Construct accessibility and depression: An examination of cognitive and affective factors. *Journal of Personality and Social Psychology*, 47(2), 427–439. <https://doi.org/10.1037/0022-3514.47.2.427>
- **Grandjean, J., D'Ostilio, K., Phillips, C., Baiteau, E., Degueldre, C., Luxen, A., Maquet, P., Salmon, E., & Collette, F. (2012). Modulation of brain activity during a stroop inhibitory task by the kind of cognitive control required. *PLoS ONE*, 7(7), e41513. <https://doi.org/10.1371/journal.pone.0041513>
- *Grandjean, J., D'Ostilio, K., Fias, W., Phillips, C., Baiteau, E., Degueldre, C., Luxen, A., Maquet, P., Salmon, E., & Collette, F. (2013). Exploration of the mechanisms underlying the ISPC effect: Evidence from behavioral and neuroimaging data. *Neuropsychologia*, 51(6), 1040–1049. <https://doi.org/10.1016/j.neuropsychologia.2013.02.015>
- Gruner, P., & Pittenger, C. (2017). Cognitive inflexibility in obsessive-compulsive disorder. *Neuroscience*, 345, 243–255. <https://doi.org/10.1016/j.neuroscience.2016.07.030>
- **Hart, S. J., Green, S. R., Casp, M., & Belger, A. (2010). Emotional priming effects during Stroop task performance. *Neuroimage*, 49(3), 2662–2670. <https://doi.org/10.1016/j.neuroimage.2009.10.076>
- *Hassel, S., Sharma, G. B., Alders, G. L., Davis, A. D., Arnott, S. R., Frey, B. N., Hall, G. B., Harris, J. K., Lam, R. W., Milev, R., Muller, D. J., Rotzinger, S., Zamyadi, M., Kennedy, S. H., Strother, S. C., & MacQueen, G. M. (2020). Reliability of a functional magnetic resonance imaging task of emotional conflict in

- healthy participants. *Human Brain Mapping*, 41(6), 1400–1415. <https://doi.org/10.1002/hbm.24883>
- Hasshim, N., & Parris, B. A. (2014). Two-to-one color-response mapping and the presence of semantic conflict in the Stroop task. *Frontiers in Psychology*, 5, 1157. <https://doi.org/10.3389/fpsyg.2014.01157>
- Hasshim, N., & Parris, B. A. (2015). Assessing stimulus-stimulus (semantic) conflict in the Stroop task using saccadic two-to-one color response mapping and prereponse pupillary measures. *Attention Perception and Psychophysics*, 77(8), 2601–2610. <https://doi.org/10.3758/s13414-015-0971-9>
- Hasshim, N., & Parris, B. A. (2018). Trial type mixing substantially reduces the response set effect in the Stroop task. *Acta Psychologica*, 189, 43–53. <https://doi.org/10.1016/j.actpsy.2017.03.002>
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1), 39–65. <https://doi.org/10.1002/jrsm.5>
- Hilbert, S., Nakagawa, T. T., Bindl, M., & Buhner, M. (2014). The spatial Stroop effect: A comparison of color-word and position-word interference. *Psychonomic Bulletin and Review*, 21(6), 1509–1515. <https://doi.org/10.3758/s13423-014-0631-4>
- **Hinault, T., Larcher, K., Zazubovits, N., Gotman, J., & Dagher, A. (2019). Spatio-temporal patterns of cognitive control revealed with simultaneous electroencephalography and functional magnetic resonance imaging. *Human Brain Mapping*, 40(1), 80–97. <https://doi.org/10.1002/hbm.24356>
- **Hoogveen, S., Snoek, L., & van Elk, M. (2020). Religious belief and cognitive conflict sensitivity: A preregistered fMRI study. *Cortex*, 129, 247–265. <https://doi.org/10.1016/j.cortex.2020.04.011>
- **Hough, C. M., Luks, T. L., Lai, K., Vigil, O., Guillory, S., Nongpiur, A., Fekri, S. M., Kupferman, E., Mathalon, D. H., & Mathews, C. A. (2016). Comparison of brain activation patterns during executive function tasks in hoarding disorder and non-hoarding OCD. *Psychiatry Research Neuroimaging*, 255, 50–59. <https://doi.org/10.1016/j.psychres.2016.07.007>
- **Huang, S., Zhu, Z., Zhang, W., Chen, Y., & Zhen, S. (2017). Trait impulsivity components correlate differently with proactive and reactive control. *PLoS ONE*, 12(4), e0176102. <https://doi.org/10.1371/journal.pone.0176102>
- Huang, Y. J., Su, L., & Ma, Q. G. (2020). The Stroop effect: An activation likelihood estimation meta-analysis in healthy young adults. *Neuroscience Letters*, 716. <https://doi.org/10.1016/j.neulet.2019.134683>
- **Huang, C. M., Polk, T. A., Goh, J. O., & Park, D. C. (2012). Both left and right posterior parietal activations contribute to compensatory processes in normal aging. *Neuropsychologia*, 50(1), 55–66. <https://doi.org/10.1016/j.neuropsychologia.2011.10.022>
- Jacoby, L. L., Lindsay, D. S., & Hessels, S. (2003). Item-specific control of automatic processes: Stroop process dissociations. *Psychonomic Bulletin and Review*, 10(3), 638–644. <https://doi.org/10.3758/bf03196526>
- **Jarcho, J. M., Fox, N. A., Pine, D. S., Etkin, A., Leibenluft, E., Shechner, T., & Ernst, M. (2013). The neural correlates of emotion-based cognitive control in adults with early childhood behavioral inhibition. *Biological Psychology*, 92(2), 306–314. <https://doi.org/10.1016/j.biopsycho.2012.09.008>
- **Jaspar, M., Genon, S., Muto, V., Meyer, C., Manard, M., Dideberg, V., Bours, V., Salmon, E., Maquet, P., & Collette, F. (2014). Modulating effect of COMT genotype on the brain regions underlying proactive control process during inhibition. *Cortex*, 50, 148–161. <https://doi.org/10.1016/j.cortex.2013.06.003>
- Jensen, A. R., & Reed, T. E. (1990). Simple reaction time as a suppressor variable in the chronometric study of intelligence. *Intelligence*, 14, 375–388. [https://doi.org/10.1016/S0160-2896\(05\)80011-X](https://doi.org/10.1016/S0160-2896(05)80011-X)
- Kalanthoff, E., Davelaar, E. J., Henik, A., Goldfarb, L., & Usher, M. (2018). Task conflict and proactive control: A computational theory of the Stroop task. *Psychological Review*, 125(1), 59–82. <https://doi.org/10.1037/rev0000083>
- **Kaufmann, L., Koppelstaetter, F., Delazer, M., Siedentopf, C., Rhomberg, P., Golaszewski, S., Felber, S., & Ischebeck, A. (2005). Neural correlates of distance and congruity effects in a numerical Stroop task: An event-related fMRI study. *Neuroimage*, 25(3), 888–898. <https://doi.org/10.1016/j.neuroimage.2004.12.041>
- Keha, E., & Kalanthoff, E. (2023). What is word? The boundary conditions of task conflict in the Stroop task. *Psychological Research*, 87(4), 1208–1218. <https://doi.org/10.1007/s00426-022-01738-z>
- **Kerns, J. G., Cohen, J. D., MacDonald, A. W., 3rd, Johnson, M. K., Stenger, V. A., Aizenstein, H., & Carter, C. S. (2005). Decreased conflict- and error-related activity in the anterior cingulate cortex in subjects with schizophrenia. *American Journal of Psychiatry*, 162(10), 1833–1839. <https://doi.org/10.1176/appi.ajp.162.10.1833>
- Kim, H. (2010). Dissociating the roles of the default-mode, dorsal, and ventral networks in episodic memory retrieval [Meta-Analysis Research Support, Non-U.S. Gov't]. *Neuroimage*, 50(4), 1648–1657. <https://doi.org/10.1016/j.neuroimage.2010.01.051>
- Kim, C., Chung, C., & Kim, J. (2012). Conflict adjustment through domain-specific multiple cognitive control mechanisms. *Brain Research*, 1444, 55–64. <https://doi.org/10.1016/j.brainres.2012.01.023>
- **Kim, C., Johnson, N. F., & Gold, B. T. (2014). Conflict adaptation in prefrontal cortex: now you see it, now you don't. *Cortex*, 50, 76–85. <https://doi.org/10.1016/j.cortex.2013.08.011>
- **Kim, C., Kroger, J. K., & Kim, J. (2011). A functional dissociation of conflict processing within anterior cingulate cortex. *Human Brain Mapping*, 32(2), 304–312. <https://doi.org/10.1002/hbm.21020>
- Koch, I., Ruge, H., Brass, M., Rubin, O., Meiran, N., & Prinz, W. (2003). Equivalence of cognitive processes in brain imaging and behavioral studies: evidence from task switching. *Neuroimage*, 20(1), 572–577. [https://doi.org/10.1016/s1053-8119\(03\)00206-4](https://doi.org/10.1016/s1053-8119(03)00206-4)
- *Kohler, S., Bar, K. J., & Wagner, G. (2016). Differential involvement of brainstem noradrenergic and midbrain dopaminergic nuclei in cognitive control. *Human Brain Mapping*, 37(6), 2305–2318. <https://doi.org/10.1002/hbm.23173>
- **Kohn, N., & Fernandez, G. (2020). Emotion and sex of facial stimuli modulate conditional automaticity in behavioral and neuronal interference in healthy men. *Neuropsychologia*, 145, 106592. <https://doi.org/10.1016/j.neuropsychologia.2017.12.001>
- Kornblum, S., & Lee, J. W. (1995). Stimulus-response compatibility with relevant and irrelevant stimulus dimensions that do and do not overlap with the response. *Journal of Experimental Psychology: Human Perception and Performance*, 21(4), 855–875. <https://doi.org/10.1037/0096-1523.21.4.855>
- Kornblum, S., & Stevens, G. (2002). Sequential effects of dimensional overlap: Findings and issues. In W. Prinz & B. Hommel (Eds.), *Attention and performance XIX: Common mechanisms in perception and action* (pp. 9–54). Oxford University Press.
- Koten, J. W., Langner, R., Wood, G., & Willmes, K. (2013). Are reaction times obtained during fMRI scanning reliable and valid measures of behavior? *Experimental Brain Research*, 227(1), 93–100. <https://doi.org/10.1007/s00221-013-3488-2>
- **Kozasa, E. H., Balarin, J. B., Sato, J. R., Chaim, K. T., Lacerda, S. S., Radvany, J., Mello, L., & Amaro, E., Jr. (2018). Effects of a 7-day meditation retreat on the brain function of meditators and non-meditators during an attention task. *Frontiers in*

- Human Neuroscience*, 12, 222. <https://doi.org/10.3389/fnhum.2018.00222>
- **Kronhaus, D. M., Lawrence, N. S., Williams, A. M., Frangou, S., Brammer, M. J., Williams, S. C., Andrew, C. M., & Phillips, M. L. (2006). Stroop performance in bipolar disorder: Further evidence for abnormalities in the ventral prefrontal cortex. *Bipolar Disord*, 8(1), 28–39. <https://doi.org/10.1111/j.1399-5618.2006.00282.x>
- **Kronke, K. M., Wolff, M., Mohr, H., Kraplin, A., Smolka, M. N., Buhlinger, G., & Goschke, T. (2018). Monitor yourself! Deficient error-related brain activity predicts real-life self-control failures. *Cognitive, Affective & Behavioral Neuroscience*, 18(4), 622–637. <https://doi.org/10.3758/s13415-018-0593-5>
- **Krug, M. K., & Carter, C. S. (2012). Proactive and reactive control during emotional interference and its relationship to trait anxiety. *Brain Research*, 1481, 13–36. <https://doi.org/10.1016/j.brainres.2012.08.045>
- *Kuhn, S., Schubert, F., Mekle, R., Wenger, E., Ittermann, B., Lindenberger, U., & Gallinat, J. (2016). Neurotransmitter changes during interference task in anterior cingulate cortex: Evidence from fMRI-guided functional MRS at 3 T. *Brain Structure & Function*, 221(5), 2541–2551. <https://doi.org/10.1007/s00429-015-1057-0>
- Lancaster, J. L., Tordesillas-Gutierrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., Mazziotta, J. C., & Fox, P. T. (2007). Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Human Brain Mapping*, 28(11), 1194–1205. <https://doi.org/10.1002/hbm.20345>
- Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin*, 139(4), 870–900. <https://doi.org/10.1037/a0030694>
- Lansbergen, M. M., Kenemans, J. L., & van Engeland, H. (2007). Stroop interference and attention-deficit/hyperactivity disorder: A review and meta-analysis. *Neuropsychology*, 21(2), 251–262. <https://doi.org/10.1037/0894-4105.21.2.251>
- Larson, G. E., Merritt, C. R., & Williams, S. E. (1988). Information processing and intelligence: Some implications of task complexity. *Intelligence*, 12, 131–147.
- **Lesh, T. A., Westphal, A. J., Niendam, T. A., Yoon, J. H., Minzenberg, M. J., Ragland, J. D., Solomon, M., & Carter, C. S. (2013). Proactive and reactive cognitive control and dorsolateral prefrontal cortex dysfunction in first episode schizophrenia. *Neuroimage. Clinical*, 2, 590–599. <https://doi.org/10.1016/j.nicl.2013.04.010>
- Leung, H. C., Skudlarski, P., Gatenby, J. C., Peterson, B. S., & Gore, J. C. (2000). An event-related functional MRI study of the Stroop color word interference task. *Cerebral Cortex*, 10(6), 552–560. <https://doi.org/10.1093/cercor/10.6.552>
- Lezak, M. D. (2012). *Neuropsychological assessment* (5th ed.). Oxford University Press.
- *Li, M., Newton, A. T., Anderson, A. W., Ding, Z., & Gore, J. C. (2019). Characterization of the hemodynamic response function in white matter tracts for event-related fMRI. *Nature Communications*, 10(1), 1140. <https://doi.org/10.1038/s41467-019-09076-2>
- Lindsay, D. S., & Jacoby, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance*, 20(2), 219–234. <https://doi.org/10.1037/0096-1523.20.2.219>
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage Publications.
- Littman, R., Keha, E., & Kalanthroff, E. (2019). Task conflict and task control: A mini-review. *Frontiers in Psychology*, 10, 1598. <https://doi.org/10.3389/fpsyg.2019.01598>
- **Loeffler, L. A. K., Satterthwaite, T. D., Habel, U., Schneider, F., Radke, S., & Derntl, B. (2019). Attention control and its emotion-specific association with cognitive emotion regulation in depression. *Brain Imaging and Behavior*, 13(6), 1766–1779. <https://doi.org/10.1007/s11682-019-00174-9>
- Lorenz, S., Weiner, K. S., Caspers, J., Mohlberg, H., Schleicher, A., Bludau, S., Eickhoff, S. B., Grill-Spector, K., Zilles, K., & Amunts, K. (2017). Two new cytoarchitectonic areas on the human mid-fusiform gyrus. *Cerebral Cortex*, 27(1), 373–385. <https://doi.org/10.1093/cercor/bhv225>
- Los, S. A. (1996). On the origin of mixing costs: Exploring information processing in pure and mixed blocks of trials. *Acta Psychologica*, 94(2), 145–188. [https://doi.org/10.1016/0001-6918\(95\)00050-X](https://doi.org/10.1016/0001-6918(95)00050-X)
- Luo, M., Mu, R., Liu, J. F., & Bai, F. H. (2020). Novel computerized psychometric tests as primary screening tools for the diagnosis of minimal hepatic encephalopathy. *World Journal of Clinical Cases*, 8(16), 3377–3389. <https://doi.org/10.12998/wjcc.v8.i16.3377>
- Macleod, C. M. (1991). Half a century of research on the Stroop effect - An integrative review. *Psychological Bulletin*, 109(2), 163–203. <https://doi.org/10.1037/0033-2909.109.2.163>
- MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, 4(10), 383–391. [https://doi.org/10.1016/S1364-6613\(00\)01530-8](https://doi.org/10.1016/S1364-6613(00)01530-8)
- MacLeod, C. M. (2005). The Stroop task in cognitive research. In Wenzel, A. & Rubin, D. C. (Eds.), *Cognitive Methods and Their Application to Clinical Research*. American Psychological Association.
- **Manard, M., Francois, S., Phillips, C., Salmon, E., & Collette, F. (2017). The neural bases of proactive and reactive control processes in normal aging. *Behavioural Brain Research*, 320, 504–516. <https://doi.org/10.1016/j.bbr.2016.10.026>
- Marini, F., Demeter, E., Roberts, K. C., Chelazzi, L., & Woldorff, M. G. (2016). Orchestrating proactive and reactive mechanisms for filtering distracting information: Brain-behavior relationships revealed by a mixed-design fMRI study. *Journal of Neuroscience*, 36(3), 988–1000. <https://doi.org/10.1523/JNEUROSCI.2966-15.2016>
- **Mathis, A., Schunck, T., Erb, G., Namer, I. J., & Luthringer, R. (2009). The effect of aging on the inhibitory function in middle-aged subjects: A functional MRI study coupled with a color-matched Stroop task. *International Journal of Geriatric Psychiatry*, 24(10), 1062–1071. <https://doi.org/10.1002/gps.2222>
- MATLAB. (2019). *version 9.7.0.1471314 (R2019b)*. In The MathWorks Inc.
- **Matthews, S. C., Paulus, M. P., Simmons, A. N., Nelesen, R. A., & Dimsdale, J. E. (2004). Functional subdivisions within anterior cingulate cortex and their relationship to autonomic nervous system function. *Neuroimage*, 22(3), 1151–1156. <https://doi.org/10.1016/j.neuroimage.2004.03.005>
- **Mead, L. A., Mayer, A. R., Bobholz, J. A., Woodley, S. J., Cunningham, J. M., Hammeke, T. A., & Rao, S. M. (2002). Neural basis of the Stroop interference task: Response competition or selective attention? *Journal of the International Neuropsychological Society*, 8(6), 735–742. <https://doi.org/10.1017/s1355617702860015>
- Meltzer, J. A., Negishi, M., & Constable, R. T. (2008). Biphasic hemodynamic responses influence deactivation and may mask activation in block-design fMRI paradigms. *Human Brain Mapping*, 29(4), 385–399. <https://doi.org/10.1002/hbm.20391>
- Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: A network model of insula function. *Brain Structure & Function*, 214(5–6), 655–667. <https://doi.org/10.1007/s00429-010-0262-0>
- *Milham, M. P., Banich, M. T., Webb, A., Barad, V., Cohen, N. J., Wszalek, T., & Kramer, A. F. (2001). The relative involvement of anterior cingulate and prefrontal cortex in attentional

- control depends on nature of conflict. *Brain Research: Cognitive Brain Research*, 12(3), 467–473. [https://doi.org/10.1016/S0926-6410\(01\)00076-3](https://doi.org/10.1016/S0926-6410(01)00076-3)
- *Mitchell, R. L. (2005). The BOLD response during Stroop task-like inhibition paradigms: Effects of task difficulty and task-relevant modality. *Brain and Cognition*, 59(1), 23–37. <https://doi.org/10.1016/j.bandc.2005.04.001>
- **Morgenroth, E., Orlov, N., Lythgoe, D. J., Stone, J. M., Barker, H., Munro, J., Eysenck, M., & Allen, P. (2019). Altered relationship between prefrontal glutamate and activation during cognitive control in people with high trait anxiety. *Cortex*, 117, 53–63. <https://doi.org/10.1016/j.cortex.2019.02.021>
- Mueller, S. T., & Piper, B. J. (2014). The psychology experiment building language (PEBL) and PEBL test battery. *Journal of Neuroscience Methods*, 222, 250–259. <https://doi.org/10.1016/j.jneumeth.2013.10.024>
- Müller, V. I., Cieslik, E. C., Laird, A. R., Fox, P. T., Radua, J., Mataix-Cols, D., Tench, C. R., Yarkoni, T., Nichols, T. E., Turkeltaub, P. E., Wager, T. D., & Eickhoff, S. B. (2018a). Ten simple rules for neuroimaging meta-analysis. *Neuroscience and Biobehavioral Reviews*, 84, 151–161. <https://doi.org/10.1016/j.neubiorev.2017.11.012>
- Müller, V. I., Höhner, Y., & Eickhoff, S. B. (2018b). Influence of task instructions and stimuli on the neural network of face processing: An ALE meta-analysis. *Cortex*, 103, 240–255. <https://doi.org/10.1016/j.cortex.2018.03.011>
- Mumford, J. A., Bissett, P. G., Jones, H. M., Shim, S., Rios, J. A. H., & Poldrack, R. A. (2023). The response time paradox in functional magnetic resonance imaging analyses. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-023-01760-0>
- *Nakao, T., Nakagawa, A., Yoshiura, T., Nakatani, E., Nabeyama, M., Yoshizato, C., Kudoh, A., Tada, K., Yoshioka, K., & Kawamoto, M. (2005). A functional MRI comparison of patients with obsessive-compulsive disorder and normal controls during a Chinese character Stroop task. *Psychiatry Research*, 139(2), 101–114. <https://doi.org/10.1016/j.psycyhresns.2004.12.004>
- Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic [Comment]. *Neuroimage*, 25(3), 653–660. <https://doi.org/10.1016/j.neuroimage.2004.12.005>
- *Ning, R. P. (2021). How language proficiency influences Stroop effect and reverse-Stroop effect: A functional magnetic resonance imaging study. *Journal of Neurolinguistics*, 60. <https://doi.org/10.1016/j.jneuroling.2021.101027>
- *Norris, D. G., Zysset, S., Mildner, T., & Wiggins, C. J. (2002). An investigation of the value of spin-echo-based fMRI using a Stroop color-word matching task and EPI at 3 T. *Neuroimage*, 15(3), 719–726. <https://doi.org/10.1006/nimg.2001.1005>
- **Ovaysikia, S., Tahir, K. A., Chan, J. L., & DeSouza, J. F. (2011). Word wins over face: Emotional Stroop effect activates the frontal cortical network. *Frontiers in Human Neuroscience*, 4, 234. <https://doi.org/10.3389/fnhum.2010.00234>
- **Overbeek, G., Gawne, T. J., Reid, M. A., Salibi, N., Kraguljac, N. V., White, D. M., & Lahti, A. C. (2019). Relationship between cortical excitation and inhibition and task-induced activation and deactivation: A combined magnetic resonance spectroscopy and functional magnetic resonance imaging study at 7T in first-episode psychosis. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 4(2), 121–130. <https://doi.org/10.1016/j.bpsc.2018.10.002>
- **Papalini, S., Michels, F., Kohn, N., Wegman, J., van Hemert, S., Roelofs, K., Arias-Vasquez, A., & Aarts, E. (2019). Stress matters: Randomized controlled trial on the effect of probiotics on neurocognition. *Neurobiology of Stress*, 10, 100141. <https://doi.org/10.1016/j.ynstr.2018.100141>
- *Pardo, J. V., Pardo, P. J., Janer, K. W., & Raichle, M. E. (1990). The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Proceedings of the National Academy of Sciences of the United States of America*, 87(1), 256–259. <https://doi.org/10.1073/pnas.87.1.256>
- **Park, I. H., Park, H. J., Chun, J. W., Kim, E. Y., & Kim, J. J. (2008). Dysfunctional modulation of emotional interference in the medial prefrontal cortex in patients with schizophrenia. *Neuroscience Letters*, 440(2), 119–124. <https://doi.org/10.1016/j.neulet.2008.05.094>
- Parris, B. A. (2014). Task conflict in the Stroop task: When Stroop interference decreases as Stroop facilitation increases in a low task conflict context. *Frontiers in Psychology*, 5, 1182. <https://doi.org/10.3389/fpsyg.2014.01182>
- Parris, B. A., Wadley, M. G., Hasshim, N., Benattayallah, A., Augustinova, M., & Ferrand, L. (2019). An fMRI study of response and semantic conflict in the Stroop task. *Frontiers in Psychology*, 10, 2426. <https://doi.org/10.3389/fpsyg.2019.02426>
- Parris, B. A., Hasshim, N., Wadley, M., Augustinova, M., & Ferrand, L. (2022). The loci of Stroop effects: A critical review of methods and evidence for levels of processing contributing to color-word Stroop effects and the implications for the loci of attentional selection. *Psychological Research*, 86(4), 1029–1053. <https://doi.org/10.1007/s00426-021-01554-x>
- Penner, I. K., Kobel, M., Stocklin, M., Weber, P., Opwis, K., & Calabrese, P. (2012). The Stroop task: Comparison between the original paradigm and computerized versions in children and adults. *Clinical Neuropsychologist*, 26(7), 1142–1153. <https://doi.org/10.1080/13854046.2012.713513>
- Perret, E. (1974). The left frontal lobe of man and the suppression of habitual responses in verbal categorical behaviour. *Neuropsychologia*, 12(3), 323–330. [https://doi.org/10.1016/0028-3932\(74\)90047-5](https://doi.org/10.1016/0028-3932(74)90047-5)
- **Peven, J. C., Litz, G. A., Brown, B., Xie, X., Grove, G. A., Watt, J. C., & Erickson, K. I. (2019). Higher cardiorespiratory fitness is associated with reduced functional brain connectivity during performance of the Stroop task. *Brain plasticity*, 5(1), 57–67. <https://doi.org/10.3233/BPL-190085>
- **Piai, V., Roelofs, A., Acheson, D. J., & Takashima, A. (2013). Attention for speaking: Domain-general control from the anterior cingulate cortex in spoken word production. *Frontiers in Human Neuroscience*, 7, 832. <https://doi.org/10.3389/fnhum.2013.00832>
- Pilli, R., Naidu, M., Pingali, U. R., Shobha, J. C., & Reddy, A. P. (2013). A computerized Stroop test for the evaluation of psychotropic drugs in healthy participants. *Indian Journal of Psychological Medicine*, 35(2), 180–189. <https://doi.org/10.4103/0253-7176.116251>
- *Pinel, P., Piazza, M., Le Bihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron*, 41(6), 983–993. [https://doi.org/10.1016/S0896-6273\(04\)00107-2](https://doi.org/10.1016/S0896-6273(04)00107-2)
- *Polk, T. A., Drake, R. M., Jonides, J. J., Smith, M. R., & Smith, E. E. (2008). Attention enhances the neural processing of relevant features and suppresses the processing of irrelevant features in humans: A functional magnetic resonance imaging study of the Stroop task. *Journal of Neuroscience*, 28(51), 13786–13792. <https://doi.org/10.1523/JNEUROSCI.1026-08.2008>
- **Pompei, F., Jogia, J., Tatarelli, R., Girardi, P., Rubia, K., Kumari, V., & Frangou, S. (2011). Familial and disease specific abnormalities in the neural correlates of the Stroop Task in Bipolar Disorder. *Neuroimage*, 56(3), 1677–1684. <https://doi.org/10.1016/j.neuroimage.2011.02.052>
- **Portes, B., Balardin, J. B., Lacerda, S., Pires, F., Tobo, P., Barichello, C., Peterson, J., Sanches, L. R., Sanches-Rocha, L., Amaro, E., Jr., & Kozasa, E. H. (2019). The effects of perceived chronic stress on the fMRI correlates of attentional control in

- women managers. *Archives of Women's Mental Health*, 22(3), 375–381. <https://doi.org/10.1007/s00737-018-0902-6>
- *Potenza, M. N., Leung, H. C., Blumberg, H. P., Peterson, B. S., Fulbright, R. K., Lacadie, C. M., Skudlarski, P., & Gore, J. C. (2003). An fMRI Stroop task study of ventromedial prefrontal cortical function in pathological gamblers. *American Journal of Psychiatry*, 160(11), 1990–1994. <https://doi.org/10.1176/appi.ajp.160.11.1990>
- **Prakash, R. S., Erickson, K. I., Colcombe, S. J., Kim, J. S., Voss, M. W., & Kramer, A. F. (2009). Age-related differences in the involvement of the prefrontal cortex in attentional control. *Brain and Cognition*, 71(3), 328–335. <https://doi.org/10.1016/j.bandc.2009.07.005>
- *Purmann, S., & Pollmann, S. (2015). Adaptation to recent conflict in the classical color-word Stroop-task mainly involves facilitation of processing of task-relevant information. *Frontiers in Human Neuroscience*, 9, 88. <https://doi.org/10.3389/fnhum.2015.00088>
- Pustejovsky, J. (2021). *Cluster-robust (sandwich) variance estimators with small-sample corrections*. <https://github.com/jepusto/clubSandwich>. Accessed Aug 2021
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>. Accessed Aug 2021
- Rabi, R., Vasquez, B. P., Alain, C., Hasher, L., Belleville, S., & Anderson, N. D. (2020). Inhibitory control deficits in individuals with amnesic mild cognitive impairment: A meta-analysis. *Neuropsychology Review*, 30(1), 97–125. <https://doi.org/10.1007/s11065-020-09428-6>
- **Ramm, M., Sundermann, B., Gomes, C. A., Moddel, G., Langenbruch, L., Nayyeri, M. D., Young, P., Pfeiderer, B., Krebs, R. M., & Axmacher, N. (2021). Probing the relevance of the hippocampus for conflict-induced memory improvement. *Neuroimage*, 226, 117563. <https://doi.org/10.1016/j.neuroimage.2020.117563>
- *Ravnikilde, B., Videbech, P., Rosenberg, R., Gjedde, A., & Gade, A. (2002). Putative tests of frontal lobe function: A PET-study of brain activation during Stroop's test and verbal fluency. *Journal of Clinical and Experimental Neuropsychology*, 24(4), 534–547. <https://doi.org/10.1076/jcen.24.4.534.1033>
- Reid, A. T., Bzdok, D., Genon, S., Langner, R., Müller, V. I., Eickhoff, C. R., Hoffstaedter, F., Cieslik, E. C., Fox, P. T., Laird, A. R., Amunts, K., Caspers, S., & Eickhoff, S. B. (2016). ANIMA: A data-sharing initiative for neuroimaging meta-analyses. *Neuroimage*, 124(Pt B), 1245–1253. <https://doi.org/10.1016/j.neuroimage.2015.07.060>
- *Roberts, K. L., & Hall, D. A. (2008). Examining a supramodal network for conflict processing: A systematic review and novel functional magnetic resonance imaging data for related visual and auditory Stroop tasks. *Journal of Cognitive Neuroscience*, 20(6), 1063–1078. <https://doi.org/10.1162/jocn.2008.20074>
- **Robertson, B. D., Hiebert, N. M., Seergobin, K. N., Owen, A. M., & MacDonald, P. A. (2015). Dorsal striatum mediates cognitive control, not cognitive effort per se, in decision-making: An event-related fMRI study. *NeuroImage*, 114, 170–184. <https://doi.org/10.1016/j.neuroimage.2015.03.082>
- Rodgers, M. A., & Pustejovsky, J. E. (2020). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods*. <https://doi.org/10.1037/met0000300>
- Rohatki, A. (2021). *Webplotdigitizer. Version 4.5*. <https://automeris.io/WebPlotDigitizer>. Accessed Nov 2021
- **Roth, R. M., Koven, N. S., Randolph, J. J., Flashman, L. A., Pixley, H. S., Ricketts, S. M., Wishart, H. A., & Saykin, A. J. (2006). Functional magnetic resonance imaging of executive control in bipolar disorder. *Neuroreport*, 17(11), 1085–1089. <https://doi.org/10.1097/01.wnr.0000227979.06013.57>
- Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., Fox, P. T., & Eickhoff, S. B. (2012). Modelling neural correlates of working memory: A coordinate-based meta-analysis. *Neuroimage*, 60(1), 830–846. <https://doi.org/10.1016/j.neuroimage.2011.11.050>
- RStudio Team. (2021). *RStudio: Integrated Development Environment for R*. In RStudio, PBC. <http://www.rstudio.com/>. Accessed Aug 2021
- **Ruff, C. C., Woodward, T. S., Laurens, K. R., & Liddle, P. F. (2001). The role of the anterior cingulate cortex in conflict processing: Evidence from reverse Stroop interference. *Neuroimage*, 14(5), 1150–1158. <https://doi.org/10.1006/nimg.2001.0893>
- *Salgado-Pineda, P., Rodriguez-Jimenez, R., Moreno-Ortega, M., Dompablo, M., Martinez de Aragon, A., Salvador, R., McKenna, P. J., Pomarol-Clotet, E., & Palomo, T. (2021). Activation and deactivation patterns in schizophrenia during performance of an fMRI adapted version of the Stroop task. *Journal of Psychiatric Research*, 144, 1–7. <https://doi.org/10.1016/j.jpsychires.2021.09.039>
- Salo, R., Henik, A., & Robertson, L. C. (2001). Interpreting Stroop interference: An analysis of differences between task versions. *Neuropsychology*, 15(4), 462–471. <https://doi.org/10.1037/0894-4105.15.4.462>
- Scheperjans, F., Eickhoff, S. B., Homke, L., Mohlberg, H., Hermann, K., Amunts, K., & Zilles, K. (2008a). Probabilistic maps, morphometry, and variability of cytoarchitectonic areas in the human superior parietal cortex. *Cerebral Cortex*, 18(9), 2141–2157. <https://doi.org/10.1093/cercor/bhm241>
- Scheperjans, F., Hermann, K., Eickhoff, S. B., Amunts, K., Schleicher, A., & Zilles, K. (2008b). Observer-independent cytoarchitectonic mapping of the human superior parietal cortex [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. *Cerebral Cortex*, 18(4), 846–867. <https://doi.org/10.1093/cercor/bhm116>
- **Schmidt, C., Peigneux, P., Leclercq, Y., Sterpenich, V., Vandewalle, G., Phillips, C., Berthomier, P., Berthomier, C., Tinguely, G., Gais, S., Schabus, M., Desseilles, M., Dang-Vu, T., Salmon, E., Degueldre, C., Baletau, E., Luxen, A., Cajochen, C., Maquet, P., & Collette, F. (2012). Circadian preference modulates the neural substrate of conflict processing across the day. *PLoS ONE*, 7(1), e29658. <https://doi.org/10.1371/journal.pone.0029658>
- **Schulte, T., Muller-Oehring, E. M., Vinco, S., Hoeft, F., Pfefferbaum, A., & Sullivan, E. V. (2009). Double dissociation between action-driven and perception-driven conflict resolution invoking anterior versus posterior brain systems. *Neuroimage*, 48(2), 381–390. <https://doi.org/10.1016/j.neuroimage.2009.06.058>
- **Schulte, T., Muller-Oehring, E. M., Sullivan, E. V., & Pfefferbaum, A. (2012). Synchrony of corticostriatal-midbrain activation enables normal inhibitory control and conflict processing in recovering alcoholic men. *Biological Psychiatry*, 71(3), 269–278. <https://doi.org/10.1016/j.biopsych.2011.10.022>
- Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., Reiss, A. L., & Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience*, 27(9), 2349–2356. <https://doi.org/10.1523/JNEUROSCI.5587-06.2007>
- *Seok Jeong, B., Kwon, J. S., Yoon Kim, S., Lee, C., Yoon, T., Moon, C. H., & Yoon Kim, C. (2005). Functional imaging evidence of the relationship between recurrent psychotic episodes and neurodegenerative course in schizophrenia. *Psychiatry Research*, 139(3), 219–228. <https://doi.org/10.1016/j.psychresns.2004.01.008>
- Shao, K., Wang, W., Guo, S. Z., Dong, F. M., Yang, Y. M., Zhao, Z. M., Jia, Y. L., & Wang, J. H. (2020). Assessing executive function following the early stage of mild Ischemic stroke with three brief screening tests. *Journal of Stroke and Cerebrovascular Diseases*,

- 29(8), 104960. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2020.104960>
- Shashidhara, S., Mitchell, D. J., Erez, Y., & Duncan, J. (2019). Progressive recruitment of the frontoparietal multiple-demand system with increased task complexity, time pressure, and reward. *Journal of Cognitive Neuroscience*, 31(11), 1617–1630. https://doi.org/10.1162/jocn_a_01440
- **Shashidhara, S., Spronkers, F. S., & Erez, Y. (2020). Individual-subject functional localization increases univariate activation but not multivariate pattern discriminability in the “multiple-demand” frontoparietal network. *Journal of Cognitive Neuroscience*, 32(7), 1348–1368. https://doi.org/10.1162/jocn_a_01554
- *Sheu, L. K., Jennings, J. R., & Gianaros, P. J. (2012). Test-retest reliability of an fMRI paradigm for studies of cardiovascular reactivity. *Psychophysiology*, 49(7), 873–884. <https://doi.org/10.1111/j.1469-8986.2012.01382.x>
- Shichel, I., & Tzelgov, J. (2018). Modulation of conflicts in the Stroop effect. *Acta Psychologica*, 189, 93–102. <https://doi.org/10.1016/j.actpsy.2017.10.007>
- **Shin, G., & Kim, C. (2015). Neural correlates of cognitive style and flexible cognitive control. *NeuroImage*, 113, 78–85. <https://doi.org/10.1016/j.neuroimage.2015.03.046>
- *Silton, R. L., Heller, W., Towers, D. N., Engels, A. S., Spielberg, J. M., Edgar, J. C., Sass, S. M., Stewart, J. L., Sutton, B. P., Banich, M. T., & Miller, G. A. (2010). The time course of activity in dorsolateral prefrontal cortex and anterior cingulate cortex during top-down attentional control. *NeuroImage*, 50(3), 1292–1302. <https://doi.org/10.1016/j.neuroimage.2009.12.061>
- **Song, Y., & Hakoda, Y. (2015). An fMRI study of the functional mechanisms of Stroop/reverse-Stroop effects. *Behavioural Brain Research*, 290, 187–196. <https://doi.org/10.1016/j.bbr.2015.04.047>
- Song, S., Zilverstand, A., Song, H., d’Oleire Uquillas, F., Wang, Y., Xie, C., Cheng, L., & Zou, Z. (2017). The influence of emotional interference on cognitive control: A meta-analysis of neuroimaging studies using the emotional Stroop task. *Scientific Reports*, 7(1), 2088. <https://doi.org/10.1038/s41598-017-02266-2>
- Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer’s type. *Journal of Experimental Psychology: Human Perception and Performance*, 22(2), 461–479. <https://doi.org/10.1037/0096-1523.22.2.461>
- Sridharan, D., Levitin, D. J., & Menon, V. (2008). A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proceedings of the National Academy of Sciences of the United States of America*, 105(34), 12569–12574. <https://doi.org/10.1073/pnas.0800005105>
- *Steel, C., Haworth, E. J., Peters, E., Hemsley, D. R., Sharma, T., Gray, J. A., Pickering, A., Gregory, L., Simmons, A., Bullmore, E. T., & Williams, S. C. (2001). *Neuroimaging Correlates of Negative Priming*. *Neuroreport*, 12(16), 3619–3624. <https://doi.org/10.1097/00001756-200111160-00049>
- Stenberg, G., Wiking, S., & Dahl, M. (1998). Judging words at face value: Interference in a word processing task reveals automatic processing of affective facial expressions. *Cognition & Emotion*, 12(6), 755–782. <https://doi.org/10.1080/026999398379420>
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662. <https://doi.org/10.1037/h0054651>
- Sun, H., Soh, K. G., Roslan, S., Wazir, M., & Soh, K. L. (2021). Does mental fatigue affect skilled performance in athletes? A Systematic Review. *Plos One*, 16(10), e0258307. <https://doi.org/10.1371/journal.pone.0258307>
- Tang, J., Critchley, H. D., Glaser, D. E., Dolan, R. J., & Butterworth, B. (2006). Imaging informational conflict: A functional magnetic resonance imaging study of numerical Stroop. *Journal of Cognitive Neuroscience*, 18(12), 2049–2062. <https://doi.org/10.1162/jocn.2006.18.12.2049>
- Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and spss. *Research Synthesis Methods*, 5(1), 13–30. <https://doi.org/10.1002/jrsm.1091>
- Tanner-Smith, E. E., & Tipton, E. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology*, 2, 85–112. <https://doi.org/10.1007/s40865-016-0026-5>
- **Taylor, S. F., Kornblum, S., Lauber, E. J., Minoshima, S., & Koeppe, R. A. (1997). Isolation of specific interference processing in the Stroop task: PET activation studies. *NeuroImage*, 6(2), 81–92. <https://doi.org/10.1006/nimg.1997.0285>
- **Terry, D. P., Faraco, C. C., Smith, D., Diddams, M. J., Puente, A. N., & Miller, L. S. (2012). Lack of long-term fMRI differences after multiple sports-related concussions. *Brain Injury*, 26(13–14), 1684–1696. <https://doi.org/10.3109/02699052.2012.722259>
- Tipton, E. (2013). Robust variance estimation in meta-regression with binary dependent effects. *Research Synthesis Methods*, 4(2), 169–187. <https://doi.org/10.1002/jrsm.1070>
- Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, 20(3), 375–393. <https://doi.org/10.1037/met0000011>
- Tornatore, J. B., Hill, E., Laboff, J. A., & McGann, M. E. (2005). Self-administered screening for mild cognitive impairment: Initial validation of a computerized test battery. *Journal of Neuropsychiatry and Clinical Neurosciences*, 17(1), 98–105. <https://doi.org/10.1176/jnp.17.1.98>
- Turkeltaub, P. E., Eickhoff, S. B., Laird, A. R., Fox, M., Wiener, M., & Fox, P. (2012). Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses [Research Support, N.I.H., Extramural]. *Human Brain Mapping*, 33(1), 1–13. <https://doi.org/10.1002/hbm.21186>
- *van de Meerendonk, N., Rueschemeyer, S. A., & Kolk, H. H. (2013). Language comprehension interrupted: Both language errors and word degradation activate Broca’s area. *Brain and Language*, 126(3), 291–301. <https://doi.org/10.1016/j.bandl.2013.07.004>
- van Maanen, L., Forstmann, B. U., Keuken, M. C., Wagenmakers, E. J., & Heathcote, A. (2016). The impact of MRI scanner environment on perceptual decision-making. *Behavior Research Methods*, 48(1), 184–200. <https://doi.org/10.3758/s13428-015-0563-6>
- **van’t Ent, D., den Braber, A., Rotgans, E., de Geus, E. J., & de Munck, J. C. (2014). The use of fMRI to detect neural responses to cognitive interference and planning: Evidence for a contribution of task related changes in heart rate? *Journal of Neuroscience Methods*, 229, 97–107. <https://doi.org/10.1016/j.jneumeth.2014.04.013>
- **Veroude, K., Jolles, J., Croiset, G., & Krabbendam, L. (2013). Changes in neural mechanisms of cognitive control during the transition from late adolescence to young adulthood. *Developmental Cognitive Neuroscience*, 5, 63–70. <https://doi.org/10.1016/j.dcn.2012.12.002>
- **Verstynen, T. D. (2014). The organization and dynamics of corticostriatal pathways link the medial orbitofrontal cortex to future behavioral responses. *Journal of Neurophysiology*, 112(10), 2457–2469. <https://doi.org/10.1152/jn.00221.2014>
- Videbech, P., Ravnkilde, B., Kristensen, S., Egander, A., Clemmensen, K., Rasmussen, N. A., Gjedde, A., & Rosenberg, R. (2003). The Danish PET/depression project: Poor verbal fluency performance despite normal prefrontal activation in patients with major depression. *Psychiatry Research*, 123(1), 49–63. [https://doi.org/10.1016/s0925-4927\(03\)00002-7](https://doi.org/10.1016/s0925-4927(03)00002-7)
- Viechtbauer, W., & Cheung, M. W. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2), 112–125. <https://doi.org/10.1002/jrsm.11>

- Viechtbauer, W. (2010). *Meta-Analysis Package for R*. <https://www.metafor-project.org>. Accessed Aug 2021
- Vogel, S. E., Goffin, C., Bohnenberger, J., Koschnig, K., Reishofer, G., Grabner, R. H., & Ansari, D. (2017). The left intraparietal sulcus adapts to symbolic number in both the visual and auditory modalities: Evidence from fMRI. *NeuroImage*, *153*, 16–27. <https://doi.org/10.1016/j.neuroimage.2017.03.048>
- *Wagner, G., De la Cruz, F., Schachtzabel, C., Gullmar, D., Schultz, C. C., Schlosser, R. G., Bar, K. J., & Koch, K. (2015). Structural and functional dysconnectivity of the fronto-thalamic system in schizophrenia: A DCM-DTI study. *Cortex*, *66*, 35–45. <https://doi.org/10.1016/j.cortex.2015.02.004>
- **Wallentin, M., Gravholt, C. H., & Skakkebaek, A. (2015). Broca's region and visual word form area activation differ during a predictive Stroop task. *Cortex*, *73*, 257–270. <https://doi.org/10.1016/j.cortex.2015.08.023>
- Westlin, C., Theriault, J. E., Katsumi, Y., Nieto-Castanon, A., Kucyi, A., Ruf, S. F., Brown, S. M., Pavel, M., Erdogmus, D., Brooks, D. H., Quigley, K. S., Whitfield-Gabrieli, S., & Barrett, L. F. (2023). Improving the study of brain-behavior relationships by revisiting basic assumptions. *Trends in Cognitive Sciences*, *27*(3), 246–257. <https://doi.org/10.1016/j.tics.2022.12.015>
- Whalen, P. J., Bush, G., Shin, L. M., & Rauch, S. L. (2006). The emotional counting Stroop: A task for assessing emotional interference during brain imaging. *Nature Protocols*, *1*(1), 293–296. <https://doi.org/10.1038/nprot.2006.45>
- Worringer, B., Langner, R., Koch, I., Eickhoff, S. B., Eickhoff, C. R., & Binkofski, F. C. (2019). Common and distinct neural correlates of dual-tasking and task-switching: A meta-analytic review and a neuro-cognitive processing model of human multitasking. *Brain Structure & Function*, *224*(5), 1845–1869. <https://doi.org/10.1007/s00429-019-01870-4>
- **Ye, Z., & Zhou, X. (2009). Conflict control during sentence comprehension: fMRI evidence. *NeuroImage*, *48*(1), 280–290. <https://doi.org/10.1016/j.neuroimage.2009.06.032>
- Zago, L., Petit, L., Turbelin, M. R., Andersson, F., Vigneau, M., & Tzourio-Mazoyer, N. (2008). How verbal and spatial manipulation networks contribute to calculation: An fMRI study. *Neuropsychologia*, *46*(9), 2403–2414. <https://doi.org/10.1016/j.neuropsychologia.2008.03.001>
- *Zhu, Z., Feng, G., Zhang, J. X., Li, G., Li, H., & Wang, S. (2013). The role of the left prefrontal cortex in sentence-level semantic integration. *NeuroImage*, *76*, 325–331. <https://doi.org/10.1016/j.neuroimage.2013.02.060>
- **Zoccatelli, G., Beltramello, A., Alessandrini, F., Pizzini, F. B., & Tassinari, G. (2010). Word and position interference in Stroop tasks: A behavioral and fMRI study. *Experimental Brain Research*, *207*(1–2), 139–147. <https://doi.org/10.1007/s00221-010-2433-x>
- **Zysset, S., Muller, K., Lohmann, G., & von Cramon, D. Y. (2001). Color-word matching Stroop task: Separating interference and response conflict. *NeuroImage*, *13*(1), 29–36. <https://doi.org/10.1006/nimg.2000.0665>
- **Zysset, S., Schroeter, M. L., Neumann, J., & von Cramon, D. Y. (2007). Stroop interference, hemodynamic response and aging: An event-related fMRI study. *Neurobiology of Aging*, *28*(6), 937–946. <https://doi.org/10.1016/j.neurobiolaging.2006.05.008>

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