

Variance (Un)Explained: Experimental Conditions and Temporal Dependencies Explain Similarly Small Proportions of Reaction Time Variability in Linear Models of Perceptual and Cognitive Tasks

Marlou Nadine Perquin^{1, 2, 3}, Tobias Heed^{1, 4, 5, 6}, and Christoph Kayser²

¹ Biopsychology and Cognitive Neuroscience, Faculty of Psychology and Sports Science, Bielefeld University

² Cognitive Neuroscience, Faculty of Biology, Bielefeld University

³ Department of Clinical Neurosciences, University of Cambridge

⁴ Center of Excellence Cognitive Interaction Technology, Bielefeld University

⁵ Department of Psychology, University of Salzburg

⁶ Centre for Cognitive Neuroscience, University of Salzburg

Any series of sensorimotor actions shows fluctuations in speed and accuracy from repetition to repetition, even when the sensory input and motor output requirements remain identical over time. Such fluctuations are particularly prominent in reaction time (RT) series from laboratory neurocognitive tasks. Despite their omnipresent nature, trial-to-trial fluctuations remain poorly understood. Here, we systematically analyzed RT series from various neurocognitive tasks, quantifying how much of the total trial-to-trial RT variance can be explained with general linear models (GLMs) by three sources of variability that are frequently investigated in behavioral and neuroscientific research: (1) experimental conditions, employed to induce systematic patterns in variability, (2) short-term temporal dependencies such as the autocorrelation between subsequent trials, and (3) long-term temporal trends over experimental blocks and sessions. Furthermore, we examined to what extent the explained variances by these sources are shared or unique. We analyzed 1913 unique RT series from 30 different cognitive control and perception-based tasks. On average, the three sources together explained ~8%–17% of the total variance. The experimental conditions explained on average ~2.5%–3.5% but did not share explained variance with temporal dependencies. Thus, the largest part of the trial-to-trial fluctuations in RT remained unexplained by these three sources. Unexplained fluctuations may take on nonlinear forms that are not picked up by GLMs. They may also be partially attributable to observable endogenous factors, such as fluctuations in brain activity and bodily states. Still, some extent of randomness may be a feature of the neurobiological system rather than just nuisance.

Public Significance Statement

When we perform stereotyped repetitive sensorimotor actions, these nevertheless exhibit remarkable variability over time even under laboratory conditions. We show that most of this variability is not explained by linear models including two factors frequently studied in experimental psychology or cognitive neuroscience: the manipulations introduced by the experimenters and endogenous temporal dependencies that arise over multiple timescales. This implies that moment-to-moment fluctuations in typical behavioral data are largely unpredictable by linear models. Endogenous processes may also influence these fluctuations even in highly stable environments. It remains an empirical question to what extent exogenous and endogenous sources influence fluctuations nonlinearly.

Keywords: intra-individual variability, spontaneous fluctuations, endogenous and exogenous factors, measurement error

Supplemental materials: <https://doi.org/10.1037/xge0001630.supp>

This article was published Online First September 23, 2024.

Timothy Vickery served as action editor.

Marlou Nadine Perquin  <https://orcid.org/0000-0001-9791-8202>

The raw data originated from the authors' own lab and are publicly available alongside the analysis code at <https://osf.io/pbcus/>. The authors have no conflicts of interest to disclose.

Preliminary analyses of this work were presented at the Virtual Meeting of the Society for Mathematical Psychology held in 2021. A previous

version of the article has been published on Biorxiv.

The authors thank the rectorate of Bielefeld University for funding, Aline Bompas for providing valuable feedback on an earlier version of the article, and the authors from the original studies for making the data sets publicly available.

Marlou Nadine Perquin played a lead role in formal analysis, investigation, project administration, software, visualization, and writing—original draft and an equal role in conceptualization and methodology.

continued

Any sensorimotor action we repeatedly execute, such as chopping vegetables, moving to a beat, or performing a defined move in sports, is characterized by variability in speed and accuracy from repetition to repetition. Such *trial-to-trial variability* is evident in typical laboratory tasks in which these sensorimotor actions are precisely recorded: Although these tasks strictly control the environmental context with a limited set of sensory inputs and motor outputs, participants' responses nevertheless fluctuate considerably from trial to trial. An example of this is shown in [Figure 1](#) for the Flanker task ([Figure 1A](#)), a well-known paradigm for conflict processing. The series of reaction time (RT) measurements across trials for one participant (abbreviated as RT series hereafter) performing this task are highly variable ([Figure 1B](#)). Such fluctuations remain poorly understood: Most of the time, we cannot explain why a participant responds fast or slowly in a particular trial, and why they commit an error in one but not another. In other words, the predictability of individual actions over time appears highly limited. However, this predictability is important for empirical sciences in general ([Yarkoni & Westfall, 2017](#)) and also for daily life—for example, being able to predict why people sometimes respond fast (or slowly) in critical tasks such as operating air control or driving a car (e.g., [Baldwin et al., 2017](#)). Here, we systematically examine trial-to-trial variability across a range of tasks that are commonly used in cognitive neuroscience—focusing on the linear trends as estimated by the general linear model (GLM). We concentrate on two well-documented constructs that both contribute to trial-to-trial variability, but which are typically studied separately: experimental conditions, which researchers may use to cause specific patterns in behavioral data, and task-unrelated fluctuations of behavior over time. These temporal fluctuations may occur between subsequent trials (short-term) and across blocks of trials (longer-term). Here, we examine their contributions separately, allowing for an examination of three sources: experimental condition, short-term dependencies, and long-term trends.

By implementing experimental conditions, researchers aim to induce systematic differences in behavior to gain an understanding of specific neurocognitive mechanisms. However, systematic differences between conditions are often difficult to identify in individual data series. This is illustrated for the Flanker task example in [Figure 1B](#): From the RT series, it is difficult to see whether the participant responds faster in congruent than incongruent trials. Fluctuations beyond those induced by the experimental conditions are usually considered “noise” and are removed during data analysis by computing condition means or other forms of summary statistics. The experimental questions are then answered at the level of these summary statistics (e.g., “Are participants *overall* slower on incongruent than on congruent trials?”). On this level, it is easier to identify patterns—for example, most participants are on average faster on the congruent than incongruent trials in the Flanker task ([Figure 1D](#)), and the average across participants is visibly lower in the congruent condition. Visualizing the distribution of RTs for a single participant reveals high variance within conditions ([Figure 1C](#)), which is often much larger than variance between conditions. Such

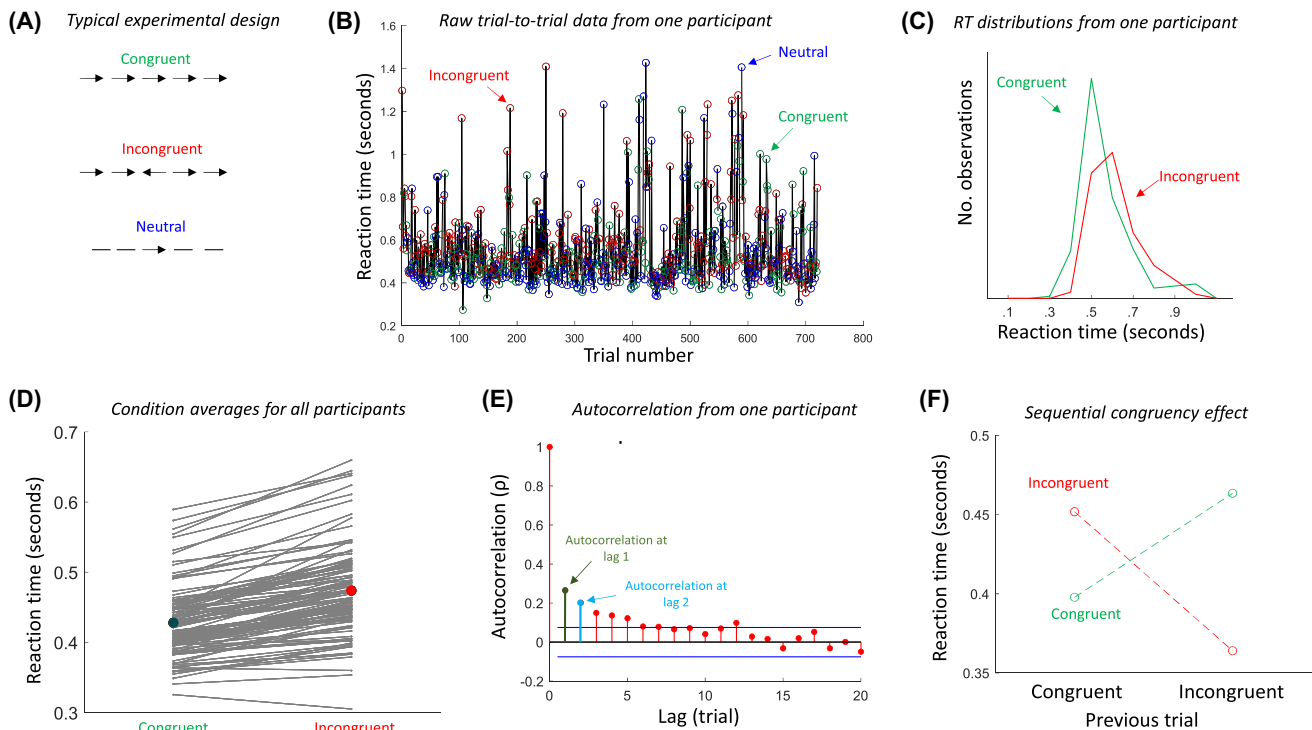
visualizations suggest that a large fraction of the within-subject variability in typical neurocognitive tasks is independent of the experimental conditions. In support of, but not directly evidencing this claim, systematic group level analyses of cognitive conflict tasks show that the between-subject variance of the difference in mean RT between congruent and incongruent conditions is on average about seven times smaller than the within-subject RT variance ([J. Rouder et al., 2019](#)). However, most studies do not quantify or report what proportion of the within-subject variance is explained by the experimental effects (but see e.g., [Williams et al., 2022](#)). To address this question, the present study offers a systematic examination of the trial-to-trial variability that is explained by the experimental conditions across a range of tasks by GLMs.

Another line of research emphasizes that the trial-by-trial variability of behavior is not random but exhibits structured dependencies over time (e.g., [Gilden, 2001](#); [Van Orden et al., 2003](#); [Perquin, van Vugt, et al., 2023](#); [Torre & Wagenmakers, 2009](#); [Torre et al., 2019](#); [Wagenmakers et al., 2004](#); [Welford, 1968](#)). For example, most RT series show positive autocorrelations, indicating that fast RTs are more likely followed by fast than slow RTs, and vice versa. Such positive autocorrelations are shown in [Figure 1E](#) for one participant in the Flanker task. Temporal dependencies have been found across many different tasks including cognitive control tasks, visual search, lexical decision, word naming, shape and color discrimination, free choice RT, and racial implicit bias tasks (e.g., [Annand & Holden, 2023](#); [Correll, 2008](#); [Gilden, 2001](#); [Holden & Rajaraman, 2012](#); [Madurski & LeBel, 2015](#); [Van Orden et al., 2003](#); [Perquin, van Vugt, et al., 2023](#); [Simola et al., 2017](#)). Importantly, they seem particularly present in tasks in which the same sensorimotor action is performed repeatedly without any changes in experimental condition, such as tapping to a fixed interval (e.g., [Chen et al., 2002](#); [Lemoine et al., 2006](#); [Madison, 2004](#); [Perquin, van Vugt, et al., 2023](#)). Because the external environment in such tasks remains unchanged, these dependencies must be driven endogenously—that is, arising from fluctuations in the internal system. Such endogenous dependencies can emerge across multiple time scales and are important for the predictability of trial-to-trial fluctuations, as high dependency implies high predictability.

Certain forms of temporal dependencies can be caused by the order of experimental conditions and are hence induced exogenously ([Duthoo et al., 2014](#); [Egner, 2017](#); [Fritzsche et al., 2017](#); [Gratton et al., 1992](#); [S. J. Kayser & Kayser, 2018](#); [Kiyonaga et al., 2017](#)). An example is the “sequential congruency” (alternatively, “Gratton”) effect, which denotes the phenomenon that participants respond faster if the previous trial was of the same condition as the current one (e.g., both congruent in the Flanker task; [Figure 1F](#)). Their interpretation is very different from endogenously-driven temporal effects: Exogenously-driven effects reflect after-effects of external stimuli, while endogenously-driven effects may reflect temporal dynamics in spontaneous neural fluctuations (e.g., [Palva et al., 2013](#); [Smit et al., 2013](#)). In the present study, we therefore included exogenous temporal effects in the form of experimental conditions, separately from short- and long-term endogenous

Tobias Heed played an equal role in conceptualization, supervision, resources, and writing—review and editing. Christoph Kayser played a supporting role in formal analysis and an equal role in conceptualization, supervision, resources, and writing—review and editing.

Correspondence concerning this article should be addressed to Marlou Nadine Perquin, Department of Clinical Neurosciences, University of Cambridge, Herchel Smith Building, Forvie Site, Robinson Way, Cambridge CB2 0SZ, United Kingdom. Email: mnp28@medschl.cam.ac.uk

Figure 1*Example of a Typical Experimental Study Using the Flanker Task*

Note. (A) Participants discriminate the direction of the middle arrow while the flankers can be identical (congruent), contradictory (incongruent), or neutral. A researcher is examining whether participants are faster in the congruent and slower in the incongruent condition. (B) Reaction time (RT) data from one participant on the Flanker task across all trials (using data from [Hedge et al., 2018](#)). This reveals large fluctuations over trials across all experimental conditions (green, red, blue). (C) RT distributions for the congruent and incongruent condition for one participant. The participant is faster in the congruent compared to the incongruent condition, but variability within each experimental condition is much larger than the average difference between experimental conditions. (D) Means on the group level for congruent and incongruent conditions, with each line representing the RT of one participant. Nearly all participants are on average faster in the congruent than in the incongruent condition, in confirmation of the initial research question. (E) Autocorrelation in RT series from one participant. Lag refers to the distance between trial numbers—for example, lag one represents the autocorrelation between trial n and trial $n-1$. (F) The effect of experimental conditions on RT in each trial is modulated by the experimental condition on the previous trial—a phenomenon known as the sequential congruency effect. See the online article for the color version of this figure.

temporal dependencies, allowing for a direct comparison of these sources of variability.

Here, we systematically examined the variance in the trial-to-trial variability of RTs explained by the experimental conditions and endogenous temporal trends across a sample of 1913 unique manual button press RT series. We focused on RT because it provides a continuous and rich measure of behavioral variability. The data were taken from publicly available resources and our own previous work and comprised tasks commonly used in the field of cognitive control (e.g., Flanker and Stroop tasks; termed “congruency-based” in the following) and perception (e.g., detection of stimuli in noisy contexts and orientation discrimination; termed “perception-based”). The data sets are typical of experimental paradigms used in psychology and cognitive neuroscience; that is, the tasks are designed to study sensory and cognitive phenomena and are not specifically tailored to increase or reduce RT variability. We quantified how much of the total RT variability is explained by (a) the experimental conditions, (b) short-range temporal dependencies between triplets of subsequent trials, and (c) long-range temporal dependencies across entire experimental blocks.

Data sets were analyzed with GLMs, which is the standard model form to study neurocognitive effects in contemporary psychology and neuroscience ([Blanca et al., 2018](#)). We chose this model form as it allows for (a) a consistent analysis pipeline across a large number of data sets—so that the presented results are comparable across all included data sets, and (b) for systematic comparison across the three sources—as nonlinear models may dilute differences between them. However, the GLM only estimates within-subject variance that can be linearly explained by the three modeled sources. All presented results should be interpreted with these analysis choices in mind. We elaborate further on potential implications in the Discussion.

Changes in behavior over the course of an experiment may or may not be related to changes in response to different conditions. For example, if a participant becomes slower over time due to increased drowsiness, it is possible that this affects difficult conditions (e.g., incongruent) more than easier ones (e.g., congruent), which would be reflected in an increase in the RT difference between conditions over time. Likewise, if a participant becomes better at responding to difficult but not to easier conditions with more practice, the

difference between conditions decreases during the experiment. Both cases would be reflected in overlapping explained variance between experimental conditions and temporal dependencies. Indeed, in vigilance studies, responses to rare but not to frequent targets get slower and less accurate over the course of an experiment (Karimi-Rouzbahani et al., 2021; Temple et al., 2000), but it remains unclear whether such trends also exist in tasks in which the conditions are balanced and (pseudo-)randomly presented. Aside from experimental conditions, one might also expect common explained variance between short- and long-range temporal dependencies; for example, both dependencies could be characterizations of the same phenomenon, such as short-term autocorrelations producing longer trends. To address these questions of overlapping variance, we systematically examined to what extent the explained variances of the three sources are shared or unique using variance partitioning.

Method

Methods were the same for the congruency-based and the perception-based tasks and are described for both below. Note that the labels “congruency-based” and “perception-based” are used as a practical way to refer to the main field and theoretical interest of the task, not to a fundamental difference of task type. This study was not preregistered.

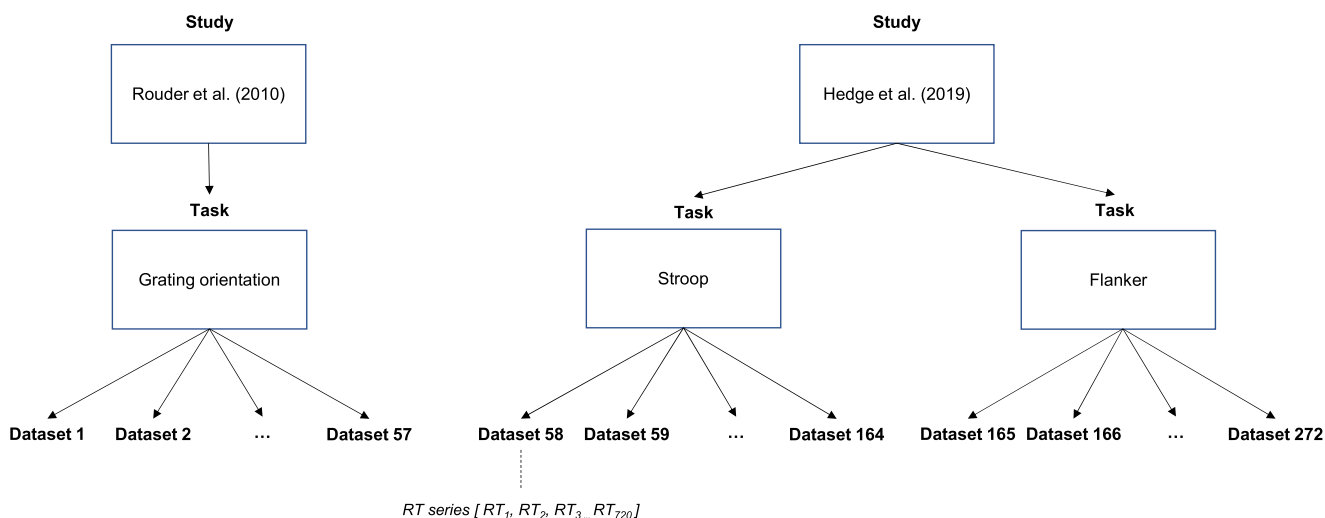
Selection of Data Sets

We collected data sets from previously published studies (Figure 2), both from publicly available sources (e.g., Open Science Framework [OSF], github) and from our own work. We applied the following criteria to select data sets:

1. The experimental conditions had to be implemented in a within-subject design as a necessary condition for studying within-subject variance.
2. Experimental conditions were presented (pseudo-) randomly across trials (rather than in a blockwise manner), allowing us to investigate the relative contributions of the sequential and naturally occurring temporal effects in experimental condition. For example, a task that requires only synchronous responses to a stimulus in one block and only asynchronous responses in the next block would not be selected.
3. The task was not explicitly intended to affect participants' performance over time; for instance, learning or fatigue-inducing studies would not meet this criterion. Such paradigms experimentally induce temporal dependencies and are therefore not of interest in the present study, as we focus on endogenous temporal dependencies.
4. The task consisted of either one or two independent “experimental manipulations of interest” to facilitate analysis, because each additional manipulation would result in an increasingly disproportional amount of predictors for a GLM. Capturing two experimental manipulations A and B in the model requires six predictors (A , B , $A \times B$, each on the current and previous trial), while three experimental conditions would require fourteen ($2 \times A$, B , C , $A \times B$, $A \times C$, $B \times C$, $A \times B \times C$), and in analogy four experimental manipulations would require 28 predictors.
5. The task consisted of minimally 20 trials for each experimental manipulation of interest, autocorrelation lag,

Figure 2

Overview of the Nomenclature in the Present Research With the Use of Two Examples



Note. “Data set” refers to the trial-to-trial data from one participant on one task, including the RT, accuracy, and experimental condition for each trial. Data sets were sourced from previously conducted studies which contained one or more tasks. For example, we sourced 57 data sets from the grating orientation task from Rouder et al. (2010), and 214 data sets across two tasks from Hedge et al. (2018). All tasks with their original source and number of data sets are shown in Tables 1 and 2. RT = reaction time. See the online article for the color version of this figure.

and block. For example, congruency-based task 1 (Table 1) has one experimental manipulation of interest (congruency), two autocorrelations, and 7 blocks of interest—which would require at least (10×20) 200 trials. Having sufficient observations per predictor is necessary for any regression model, and particularly important for the present within-subject GLMs, as RT is a measure with high variability. Though there are no agreed-upon rules for the number of observations per predictor, we relied on a guideline that is considered stringent (Ogundimu et al., 2016).

- 6. The task featured at least 15 participants to obtain a reasonable estimate of the explained variance at the task level. We put the threshold for number of participants lower than for number of trials because the analyses are run on the within-subject level.
- 7. The data set included the raw (trial-to-trial) data and, for each trial, the RT, accuracy, experimental condition(s), trial, and block number (or, if missing, can be inferred from the original article).
- 8. RT were measured using manual button presses, the most commonly used response measure.

We use the term “experimental manipulation of interest” to refer to those manipulations that were of interest in the (statistical) comparison of the task at hand. For instance, in a Stroop task, experimenters use different words and colors for the stimuli, but the interest lies in the difference between congruent, neutral, and incongruent trials independent of the specific word-color combination. As such, this task has one manipulation of interest (“congruency”) that we use throughout the main analyses, as this is the driving source of experimentally induced variance. We will address the effect of other endogenous elements in the *Control analyses* (section *Additional sources of variance*).

Selected Data Sets

Congruency-Based Data Sets

Sixteen tasks were selected for analysis (see Table 1 for an overview, and the Supplemental Materials A for a full description). Of these, eleven were collected through a previous large-scale reanalysis of congruency-based tasks (J. Rouder et al., 2019), an additional three through the Confidence Database on OSF (Rahnev et al., 2020), and two were already in possession by the first author for another study (Perquin, van Vugt, et al., 2023). We note that for these last two tasks (both from Hedge et al., 2018), the autocorrelation in RT has been previously reported alongside other temporal structure measures in Perquin, van Vugt, et al., 2023 for a study of the reliability of temporal structures. All analyzed data sets have been included in the present study.

Perception-Based Tasks

Fourteen tasks were selected for analysis (see Table 2 for an overview, and the Supplemental Materials A for a full description). Of these, six came from our own lab, one was found through the same large-scale reanalysis as above (J. Rouder et al., 2019), and an additional eight were found in the Confidence Database on OSF

Table 1
Overview of Selected Congruency-Based Tasks

	Data set	No. data set	No. excluded	Manipulation	(no. level)	Stimulus	Block	Trial/block	Total trial no.	Original article	Original exclusion rules
1. Stroop		38	1	Congruency (3)		Visual	7	72	504	Pratte et al. (2010)	<200 ms, >2000 ms
2. Simon		38	0	Congruency (3)		Visual	7	72	504	Pratte et al. (2010)	<200 ms, >2000 ms
3. Flanker		107	4	Congruency (3)		Visual	5	144	720	Hedge et al. (2018)	<100 ms, >3SD + <i>Mdn</i>
4. Stroop		107	1	Congruency (3)		Visual	5	144	720	Hedge et al. (2018)	<100 ms, >3SD + <i>Mdn</i>
5. Digit distance		54	3	Difficulty (3)		Visual	6	60	360	J. N. Rouder et al. (2005)	<250 ms, >2000 ms
6. Flanker arrow (young participants)		120	10	Congruency (3)		Visual	4	74	296	Rey-Mermet et al. (2018)	<250 ms, >2.5SD + <i>M</i>
7. Flanker arrow (older participants)		149	10	Congruency (3)		Visual	4	74	296	Rey-Mermet et al. (2018)	<250 ms, >2.5SD + <i>M</i>
8. Flanker letter (young)		120	13	Congruency (3)		Visual	4	74	296	Rey-Mermet et al. (2018)	<250 ms, >2.5SD + <i>M</i>
9. Flanker letter (old)		148	18	Congruency (3)		Visual	4	74	296	Rey-Mermet et al. (2018)	<250 ms, >2.5SD + <i>M</i>
10. Stroop color (young)		120	5	Congruency (3)		Visual	4	74	296	Rey-Mermet et al. (2018)	<250 ms, >2.5SD + <i>M</i>
11. Stroop color (old)		149	16	Congruency (3)		Visual	4	74	296	Rey-Mermet et al. (2018)	<250 ms, >2.5SD + <i>M</i>
12. Stroop number (young)		119	2	Congruency (3)		Visual	4	74	296	Rey-Mermet et al. (2018)	<250 ms, >2.5SD + <i>M</i>
13. Stroop number (old)		147	17	Congruency (3)		Visual	4	74	296	Rey-Mermet et al. (2018)	<250 ms, >2.5SD + <i>M</i>
14. Priming congruency		31	4	Congruency (2)		Visual	8	80	640	Desender et al. (2016)	N/A
15. Priming congruency		86	10	Congruency (2)		Visual	8	60	420	Desender et al. (2014)	none
16. Priming congruency		30	4	Congruency (3)		Visual	6	144	846	Teuchies et al. (2019)	>1,500 ms

Note. Overview showing the task name (Data sets), total number of participants with complete data sets (No. ppt), number of excluded participants in analysis based on % excluded trials (No. excluded; see Methods above), the experimental manipulation(s) of interest with the number of conditions in brackets (No. levels), the perceptual type of the stimulus (stimulus), the total number of blocks (blocks), the number of trials per block (trials/block) and per experiment (total trial no.), the number of independent factors to predict RT (No. predictors) for each participant, the original study in which the data was first published, and the original exclusion rules for RT outliers in the original study. RT = reaction time.

Table 2
Overview of Selected Perception-Based Data Sets

Data set	No. data set	No. excluded	Manipulation (no. level)	Stimulus	Block	Trial/block	Total trial no.	Original article	Original exclusion rules
1. Grating orientation discrimination ^a	57	5	Difficulty (3) × Direction (2)	Visual	9	80	720	J. N. Rouder et al. (2010)	<200 ms, >5,000 ms
2. Random dots direction discrimination (young) ^a	18	0	Congruency (2) × Coherence (adaptive)	Visual + Auditory	5	240	1,200	S. J. Kayser and Kayser (2018)	<300 ms, >1,200 ms
3. Random dots direction discrimination (old) ^a	23	0	Congruency (2) × Coherence (adaptive)	Visual + Auditory	3	240	720	H. Park et al. (2021)	<300 ms, >1,200 ms
4. Frequency discrimination	16	0	Frequency (7) × Delay (6)	Auditory	4	168	672	S. J. Kayser et al. (2016)	N/A
5. Intensity discrimination	16	0	Intensity (7) × Delay (6)	Auditory	Subject-dependent, ranging 3–5	168	504, 672, or 840	S. J. Kayser et al. (2016)	N/A
6. Location discrimination ^a	30	0	Modality of Stimulus (3) × Direction (2)	Visual, auditory, or visuo-auditory	4	Subject-dependent, 239 or 215	864 or 956	C. Kayser (2018)	N/A
7. Two choice noisy target detection	30	7	Location (2)	Visual	10	100	1,000	Maniscalco et al. (2017)	N/A
8. Two choice noisy target detection ^a	41	4	Location (2) × Coherence (3)	Visual	5	102	510	Maniscalco et al. (2017)	N/A
9. Grating orientation discrimination	15	1	Contrast (2)	Visual	2	160	320	Samaha et al. (2016)	<3SD + M, >3SD + M
10. Grating orientation discrimination ^a	20	0	Contrast (adaptive)	Visual	3	100	300	Samaha & Postle, 2017	<200 ms, highest 5%
11. Grating contrast discrimination ^a	19	0	Contrast (adaptive)	Visual	3	100	300	Samaha & Postle, 2017	<200 ms, highest 5%
12. Short term memory of grating orientation ^a	20	5	Contrast (adaptive)	Visual	3	100	300	Samaha & Postle, 2017	<200 ms, highest 5%
13. Color judgment	24	9	Combination of mean and variance of color (4)	Visual	8	Subject-dependent, 60 or 64	480 or 512	Desender et al. (2019)	N/A
14. Facial emotion recognition	21	1	Emotion category (2)	Visual	2	200	400	Johannknecht and Kayser (2022)	< 200 ms >3SD + M of log(RT)

Note. Same conventions as Table 1. RT = reaction time.

^a Predictors in the tasks showed signs of severe multicollinearity. Control analyses without these data sets are reported in the Supplemental Materials C.

(Rahnev et al., 2020). One additional analyzed task has not been included in the current results, as this concerned a simple detection task with a high number of anticipations, and all participants would have too little trials for the present purposes after data exclusion.

Data Preprocessing

Data cleaning was kept to a minimum to preserve within-subject fluctuations as faithfully as possible. Some cleaning is necessary: Though the origins of “outliers” remain largely unknown, they have a disproportionate weight on the fit of regression coefficients and were therefore excluded. First, excluded trials with $RT > 10$ s and $RT < 150$ ms were excluded under the assumption that these are not responses to the stimulus (but rather anticipations, experimenter errors, etc.). Next, trials with $RT > 3$ standard deviations above the mean were excluded separately for each participant and condition. The remaining RT series were inverse-transformed for all analyses to reduce their distribution’s skewness. The impact of these preprocessing choices on the overall results is discussed below (see section *Control analyses confirm that results are robust in Results*).

Trials with missing values were excluded from all analyses conducted in the present study. This seems to be the most common approach in studies on temporal dependencies, but it is important to note that the measurement of temporal dependencies is affected by trial exclusion (Karalunas et al., 2013, 2014; Kofler et al., 2013; Perquin, van Vugt, et al., 2023). Methods that deal with missing data, like data imputation, assume that data is missing at random (Donders et al., 2006). However, this is typically not the case for missing values in RT series, as these reflect moments in which performance on the task was particularly poor—and as such, there are no obvious ways to deal with missing data (see Perquin, van Vugt, et al., 2023 for a more elaborate discussion). For the current purposes, we want to avoid imputed data to cause under- or over-estimation of explained variance, and therefore exclude the respective trials.

For any missing or excluded trials, the subsequent two trials were also excluded to ensure the autocorrelations at lags 1 and 2 are only calculated on trials that directly follow each other. For the blockwise dependencies, trial indices were not altered after trial exclusion to ensure the long-term structure was preserved as much as possible. Participants were excluded from all analyses if the number of missing RTs in the full model exceeded 10%—note that this value includes both the nonresponses and excluded outliers, plus the two subsequent trials.

Data Analysis

Fitting General Linear Models

Full Model. To quantify the total explained within-subject variance, we fit GLMs for each data set on the RTs as dependent variable, using the *fitglm* function in Matlab 9 (MathWorks, 2021) with an Iterative Reweighted least squares algorithm. All GLMs included the following predictors:

1. Experimental condition(s) in the current trial,
 - a. if the design included two experimental manipulations of interest: the interaction between experimental conditions in the current trial,

2. experimental condition(s) of the previous trial,
 - a. If the design included two experimental manipulations of interest: the interaction between experimental conditions in the previous trial,
3. RT of the previous trial,
4. RT of two trials ago,
5. the linear and quadratic trends in trial number, separately for each block,
6. the linear and quadratic trends in block number,
7. the interactions between trial and block number, and
8. the interactions between trial number, block number and the experimental condition(s)—that is, Condition \times Trial \times Block, including all lower-order interactions.

Points 1 and 2 represent the experimental conditions; 3 and 4 the autocorrelation; 5, 6, and 7 the long-range trends; and 8 the interactions between experimental conditions and trends (see [Supplemental Materials A](#) for the full equations).

The experimental conditions were added as categorical variables unless specified otherwise in [Tables 1](#) and [2](#). Note that by using blockwise trial numbers, a new block marks a “reset” (e.g., after completing a previous block, having a break, or any other activities).

From the GLM fits, we extracted the amount of explained variance (R_{abc}^2), calculated as [Equation 1](#):

$$R_{abc}^2 = 1 - \frac{\text{Sum of Squares}_{\text{residual}}}{\text{Sum of Squares}_{\text{total}}} \quad (1)$$

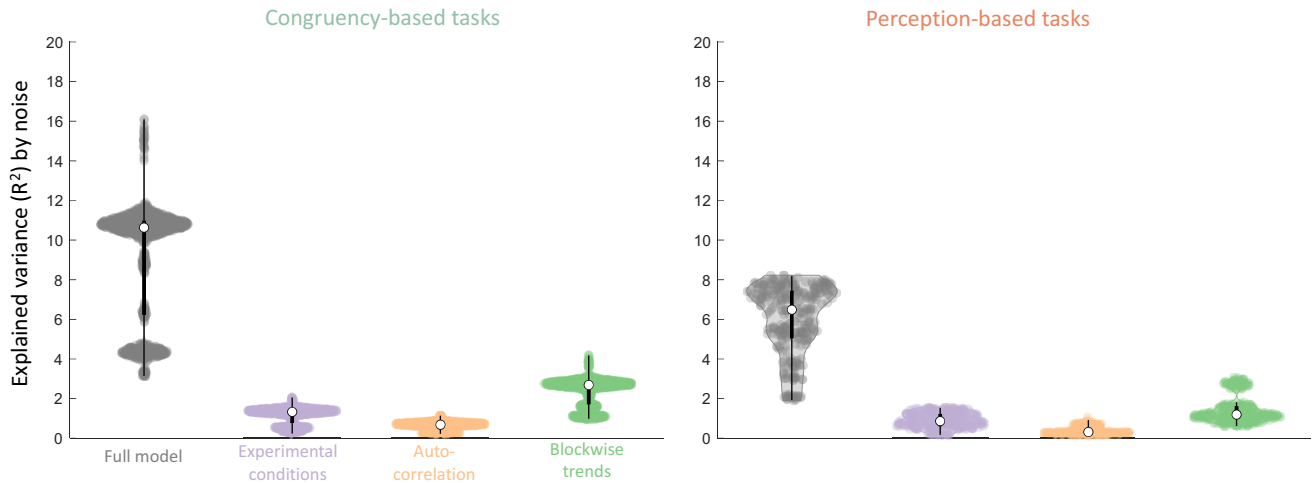
R_{abc}^2 refers to the total amount of trial-to-trial variance in RT explained by all predictors— that is, the combination of the experimental conditions (c), autocorrelation (a), and blockwise linear trend (b) all together. Hence, the amount of unexplained variance can be calculated by $1 - R_{abc}^2$.

Models that have a low number of observations, large number of to-be-estimated parameters, correlated predictors, or small effect sizes may lead to overfitting, which in turn leads to inflated estimates of explained variance (Yarkoni & Westfall, 2017). Therefore, we implemented a correction procedure separately for each data set. This was done by randomly shuffling the RT series 100 times and fitting the GLMs on each of these shuffled series, and subsequently calculating the average explained variance across these 100 iterations. This procedure results in an estimate of how much each of the predictors can explain on a noise series with the same mean and variance of the RT series. If there is no overfitting at all, these estimates of explained variance should be zero. Instead, the distributions were clearly above zero (see [Figure 3](#)—note that the estimations have been multiplied by 100 for interpretability). To correct for this bias, the mean estimate of explained variance on noise was subtracted from R_{abc}^2 for each data set. Note that these corrected variance estimates can in principle be negative, with values near zero indicating that the variance explained in the actual data is comparable to an amount of variance explained by chance.

Fitting Each Source of Variability Separately. Aside from the total amount of explained variance, we were also interested in how much variance each of the three sources—experimental

Figure 3

Estimates of Explained Variance in Percentages by GLMs on Shuffled RT Series for the Congruency-Based (Left) and Perception-Based (Right) Tasks



Note. Each dot represents one data set, the white circle represents the median of the distribution, and the width of the dot cloud indicates distribution density—by the three sources combined (full model or R^2_{abc} ; grey) as well as separately (purple, orange, and green for experimental condition R^2_c , autocorrelational R^2_a , and blockwise trends R^2_b sources respectively). Theoretically, one would expect these estimates to be zero. The distributions are clearly above zero, indicating inflation of explained variance. RT = reaction time; GLM = general linear model. See the online article for the color version of this figure.

conditions, autocorrelation, and blockwise trends—could explain by itself. We fitted separate GLMs by removing all other terms from the overall regression equation and extracted explained variance for each — R^2_c , R^2_a , and R^2_b , respectively. These three estimates were corrected by noise in the same manner as above.

These R^2 values reflect the variance each source explains in total, while not controlling for the variance it shares with the other two sources. The variance explained by the full model (R^2_{abc}) can therefore be lower than the sum of R^2_c , R^2_a , and R^2_b . This is illustrated in Figure 4C: part of the explained variance may overlap between sources (i.e., overlapping parts of the circles). To quantify the amount of variance that is *uniquely* explained by each source (i.e., the size of the non-overlapping parts of the circles), we subsequently used variance partitioning analysis (Borcard et al., 1992).

Partitioning Variance

Unique Explained Variance. To partition the variance, we first fitted each two-way combination of the three sources (experimental conditions plus autocorrelation, experimental conditions plus blockwise trends, and autocorrelation plus blockwise trends) and extracted their explained variance: R^2_{ac} , R^2_{bc} , and R^2_{ab} , respectively. These estimates were again corrected for noise as above. Next, to quantify the unique explained variance for each source, the R^2 from the other two sources combined was subtracted from the total amount of explained variance by all three sources together. For example, we calculated the unique explained variance by the experimental conditions ($R^2_{c.ab}$) as Equation 2:

$$R^2_{c.ab} = R^2_{abc} - R^2_{ab}. \quad (2)$$

Communalities. Besides the unique explained variance, variance partitioning (Borcard et al., 1992) offers a quantification of the

amount of shared variance (i.e., the overlapping parts of the circle; Figure 4C). For example, the communality C_{ab} between autocorrelation and blockwise trends (overlap in green and orange circle in Figure 4C) was calculated as Equation 3:

$$C_{ab} = R^2_{bc} + R^2_{ac} - R^2_c - R^2_{abc}, \quad (3)$$

and the communalities C_{abc} between all three sources as Equation 4:

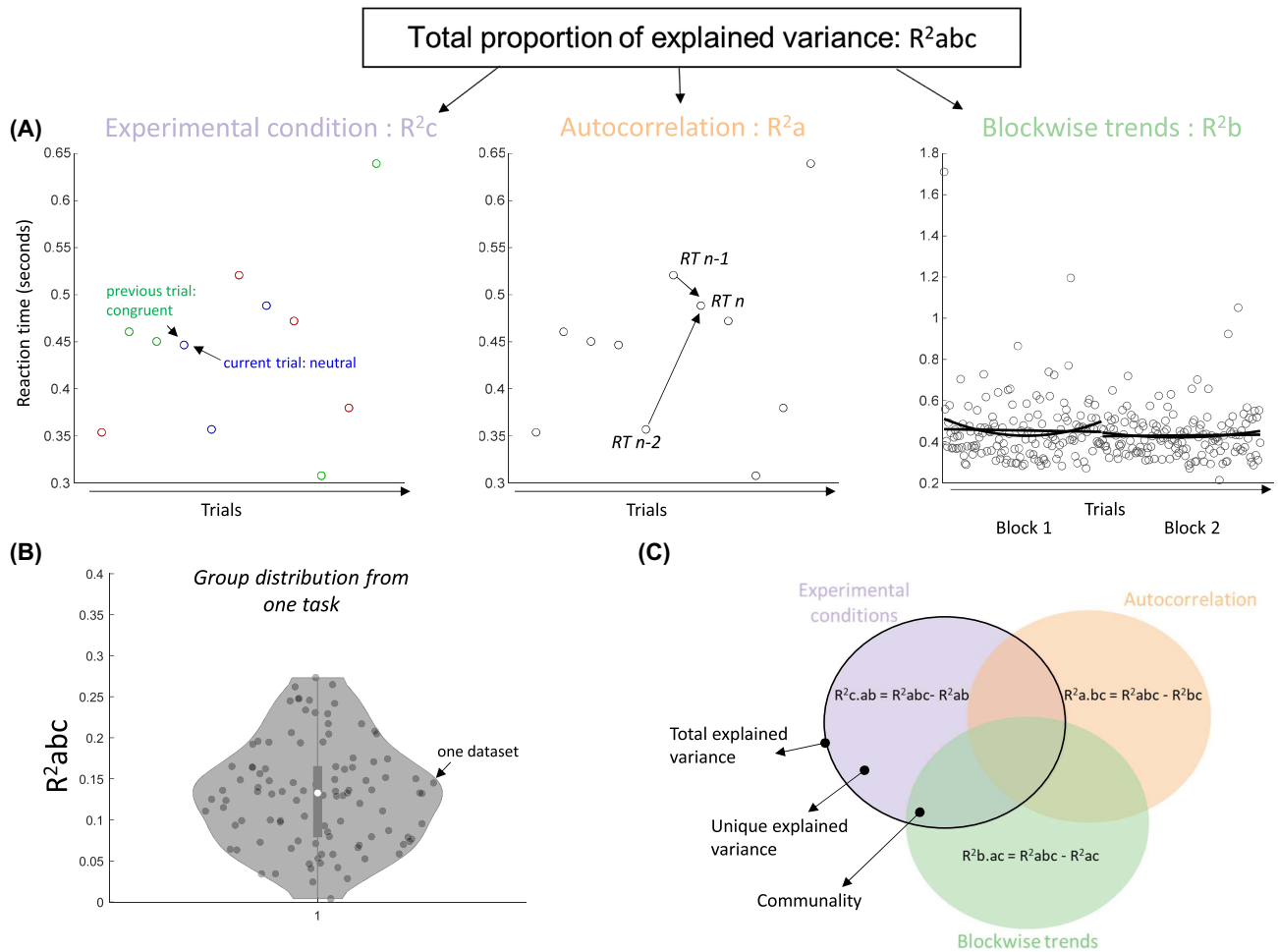
$$C_{abc} = R^2_a + R^2_b + R^2_c - R^2_{ac} - R^2_{ab} - R^2_{bc} + R^2_{abc}. \quad (4)$$

Note that theoretically, communalities cannot be < 0 . However, for the noise-corrected variance estimates used here they can be, again with values near zero indicating no common variance.

Individual Differences

To anticipate on the results, the amount of explained variance showed large differences between individuals (see spread of data points in Figure 5). We tested whether individual differences in explained variance can be explained by differences in task performance, total RT variance, or speed–accuracy trade-offs. Three performance measures were calculated for each data set: %-correct and M and SD of RT on correct trials. Second, the EZ-diffusion model was used to estimate speed–accuracy trade-offs (Wagenmakers et al., 2007). This model distinguishes strategy adjustments from true performance improvements. It is based on the drift–diffusion model by Ratcliff (1978), which aims to capture different components of decision-making into different components. The EZ-diffusion model uses calculations rather than fitting procedures and is particularly suited for identifying drift rate and boundary separation effects (van Ravenzwaaij & Oberauer, 2009; van Ravenzwaaij et al., 2017).

Figure 4
Outline of the Data Analysis



Note. (A) For each participant, GLMs were fit on RT for each of the three sources of interest: (1) experimental conditions (including the current and previous trial as predictors; left), (2) autocorrelation (including the previous and second-to-last trial; middle), and (3) linear and quadratic blockwise trends (including blockwise trial numbers as predictors; right) to extract the explained variance (R^2) for each source. Furthermore, we fit a GLM that combined these individual predictors ("full model") to extract the variance explained by all three sources together (R^2_{abc}). (B) The total proportion of explained variance was extracted for each participant for each of the sources separately (R^2_a , R^2_b , R^2_c), as well as for all three sources combined (R^2_{abc}). Distributions of explained variance are examined on the group level, showing R^2_{abc} as an example for one task. (C) The unique explained variance for each source of interest is calculated by subtracting the explained variance of the other two components from the overall explained variance. The communality in turn refers to the explained variance that is shared by each combination of sources. RT = reaction time; GLM = general linear model. See the online article for the color version of this figure.

It provides three key parameters: (1) drift rate (v) reflecting information processing speed, (2) boundary separation (α) reflecting the threshold to make a decision, and (3) nondecision time (T_{er}) reflecting perceptual and motor processes unrelated to the actual decision-making. Enhanced performance is linked to higher drift rates and shorter nondecision times, while differences in speed-accuracy trade-offs are associated with boundary separation.

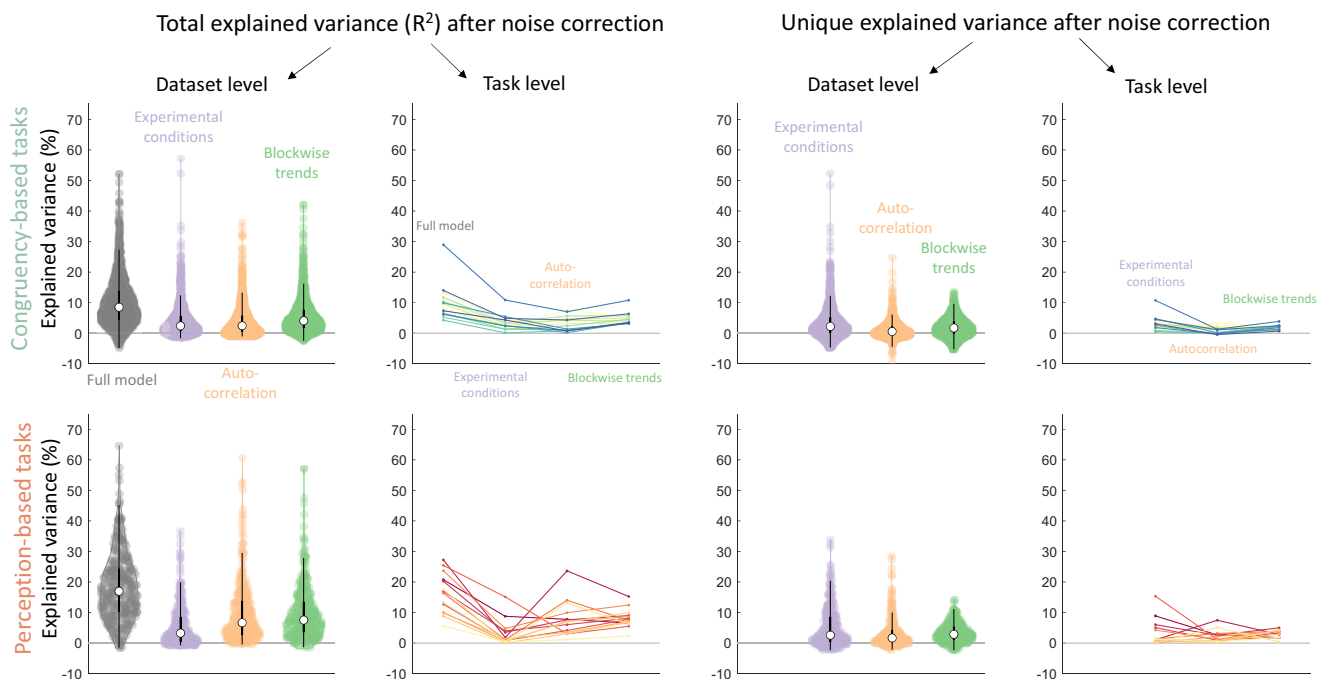
Separately for each of the 30 tasks, between-subject Pearson correlation coefficients were calculated between performance (%-correct, M and SD of RT on correct trials), drift rate, boundary separation, and explained variance (total explained variance by all sources, and unique explained variance by each source separately).

In total, this resulted in 24 correlation coefficients per data set. We used Bayesian One-Sample t -tests to compare the distributions of correlations against 0. These were calculated in JASP (JASP Team, 2017), using equal prior probabilities for each model and 10,000 Monte Carlo simulation iterations. As we did not have directional expectations, t -tests were performed two-sided.

We also examined the effect of the total number of trials for each data set. The total number of trials was correlated with explained variance (total explained variance by all sources, and unique explained variance by each source separately) across all data sets. Because the distribution of trial number was not normal, Kendall's τ correlations were computed.

Figure 5

Total Percentage of Explained Variance and Unique Explained Variance in Congruency- and Perception-Based Tasks



Note. Total percentage of explained variance (left panels) and unique explained variance (right panels) in congruency-based tasks (top rows) and perception-based tasks (bottom rows) after noise correction. On the data set level, each dot represents one data set, the white circle represents the median of the distribution, and the width of the dot cloud indicates distribution density—by the three sources combined (full model or R_{abc}^2 ; grey) as well as separately (purple, orange, and green for experimental condition R_c^2 , autocorrelational R_a^2 , and blockwise trends R_b^2 sources, respectively) as estimated by the data set level GLMs. On the task level, each colored line represents the median across data sets for each task separately across the three sources (congruency-based tasks in warm colors and perception-based tasks in cold colors). Negative values indicate instances in which the predictors explained less variance than random noise. GLMs = general linear models. See the online article for the color version of this figure.

Transparency and Openness

The data sets of experiments from our labs have been made available on Open Science Framework at <https://osf.io/pbcus/> alongside the analysis code (Perquin, Heed, & Christoph, 2023). All data from other researchers have been cited in Tables 1 and 2, as well as in the Supplemental Materials. The OSF project associated with this publication contains links to these openly available data sets from other researchers, and all original articles have been cited in Tables 1 and 2 as well as Supplemental Materials A. No original research materials have been created for the present study. The study and analysis plans were not preregistered.

Results

We analyzed RT time series from a total of 30 tasks. Sixteen of them were from congruency-based tasks (1,563 data sets in total, $N = 1,445$ after exclusion; Table 2), and 14 were from perception-based tasks (350 data sets, $N = 318$ after exclusion; Table 3).

The main analysis centered around three sources of trial-to-trial variability, with each source consisting of multiple predictors (Figure 4A). The first source is the “experimental conditions,” operationalized as the condition of the current and previous trial to capture the immediate experimental effects and the sequential condition effects. Figure 4A (left panel) shows an example for one

trial (left panel) in which the RT_n is predicted by the condition on the current trial n (neutral) and the previous trial (congruent). The second source is the “autocorrelation,” operationalized as the autocorrelation at lag 1 (= previous trial) and lag 2 (= two trials back) to capture short-range dependencies—that is, every trial n ’s RT is predicted from previous trials $n - 1$ and $n - 2$ (middle panel). The third source is the “blockwise trends,” operationalized as trial-wise linear and quadratic trends across the experimental blocks to capture long-range dependency (right panel). Because our main measure is explained variance, we emphasize the numerical values as a measure of effect sizes: Higher values indicate more explanatory power. No inferential statistics are used to compare the distributions. We corrected variance estimates by subtracting mean explained variance obtained from randomly shuffled data across 100 iterations per data set (see Methods). If there were no systematic patterns in the data by either experimental conditions or temporal dependencies, the explained variance should not be higher than the explained variance explained by chance.

To calculate the proportion of explained variance, Generalized Linear Models were fit separately for each data set, regressing RT as dependent variable. The total proportions of explained variance were extracted for each source of variance separately, as well as for all three sources combined. For example, Figure 4B shows the distribution of total explained variance for the Flanker task from

Table 3

Total Percentage of Explained Variance and Unique Explained Variance in Congruency- and Perception-Based Tasks

Source	Total % of explained within-subject variance after noise correction	% of unique explained within-subject variance
Congruency-based		
All sources	8.48 (9.2)	
Experimental condition	2.40 (4.7)	2.19 (4.9)
Autocorrelation	2.42 (5.1)	0.006 (2.7)
Blockwise trends	4.06 (5.8)	1.68 (3.9)
Perception-based		
All sources	16.98 (14.0)	
Experimental condition	3.27 (7.9)	2.57 (8.1)
Autocorrelation	6.64 (11.0)	1.68 (3.9)
Blockwise trends	7.56 (9.7)	2.88 (4.2)

Note. Median percentages of total and unique explained within-subject variance plus interquartile range in brackets on the data set level after noise correction (corresponding to the white dots in the violin plots on Figure 5). Highest values of explained variance across the three sources are displayed in bold.

Hedge et al. (2018), with each dot representing one data set and the white dot showing the median value across all data sets. Next, the unique and common proportions of explained variance were calculated for each source of variance. A visual aid for the unique and common explained variance is shown in Figure 4C in the form of Venn diagrams. For interpretation purposes, R^2 -values, as well as communalities were multiplied by 100 to obtain the percentage of variance explained.

Explained Within-Subject Variance Is Low

Figure 5 shows the explained variance for all sources combined, and each source considered individually, separately for congruency- (top rows) and perception-based (bottom) tasks. Each data set is represented by one dot in the violin (data set level)—that is, 1,763 estimates across two types of tasks of how much trial-to-trial variance is explained by the three sources. The density of the violin shows where most data sets are located, and the white dot represents the median value across all data sets. The numerical values of these medians are provided in Table 3. The total proportion of variance (left panels) explained by the experimental conditions was less than a third of the total explained variance; in other words, temporal dependencies explained more variance than experimental conditions. Moreover, the total contribution of experimental conditions was <10% for 1,313 out of all 1,445 analyzed RT series (91%) in the congruency-based tasks, and for 255 out of 318 RT series (80%) in the perception-based tasks. Numerically, blockwise trends explained the highest proportion of variance. The unique variance (right panels) explained by autocorrelation and blockwise trends were numerically substantially lower compared to their total explained variances (see Table 3). This reduction was only minimal for the experimental conditions. Numerically, median explained variance was higher for the perception- compared to the congruency-based tasks.

It may be possible that there are differences between the different tasks that would skew the data set-level distributions. We therefore calculated the group medians separately for each task to examine

consistency across tasks. These are shown in the task level panels in Figure 5, with the medians for each task being represented by one line. Violin plots across participants separately for each task can be found in Supplemental Materials B. On the task level, the three sources collectively explained on average 1/7 to 1/10 of the total trial-to-trial variance. Thus, the largest part of the variance remained unexplained when summarizing the results across tasks.

Short- and Long-Term Dependency, but Not Condition, Share Explained Variance

We also extracted the amount of explained within-subject variance that was shared between sources (i.e., communality; overlapping parts of the circles in Figure 2C). There was virtually no overlap between the variance explained by the experimental conditions and the temporal dependencies (see Figure 6 for distributions). For example, the variance commonly explained by experimental conditions and autocorrelation (blue) was on average .02% of the total within-subject variance. The other communalities involving experimental condition (grey and pink) were also negligible, both in the congruency- and perception-based tasks. The median communality value between autocorrelation and blockwise trends was 3.98% and 1.92% for congruency- and perception-based tasks, respectively. This indicates that shared explained variance between short- and long-term dependencies amounts to about 2%–4% in the total trial-to-trial variance. Overall, explained variance originating from experimental conditions was thus largely unique, whereas the two temporal sources shared explained variance.

Effect of Accuracy

RT distributions of correct and incorrect trials are known to differ both in mean and variance. We therefore ran a model to estimate the contribution of accuracy to trial-to-trial variability (Figure 7). Explained variance was increased with 1.1% by the congruency-based tasks and with 1.8% by the perception-based tasks (dark purple violins). The increase in explained variance differed noticeably across tasks—especially in the congruency-based tasks (right panels).

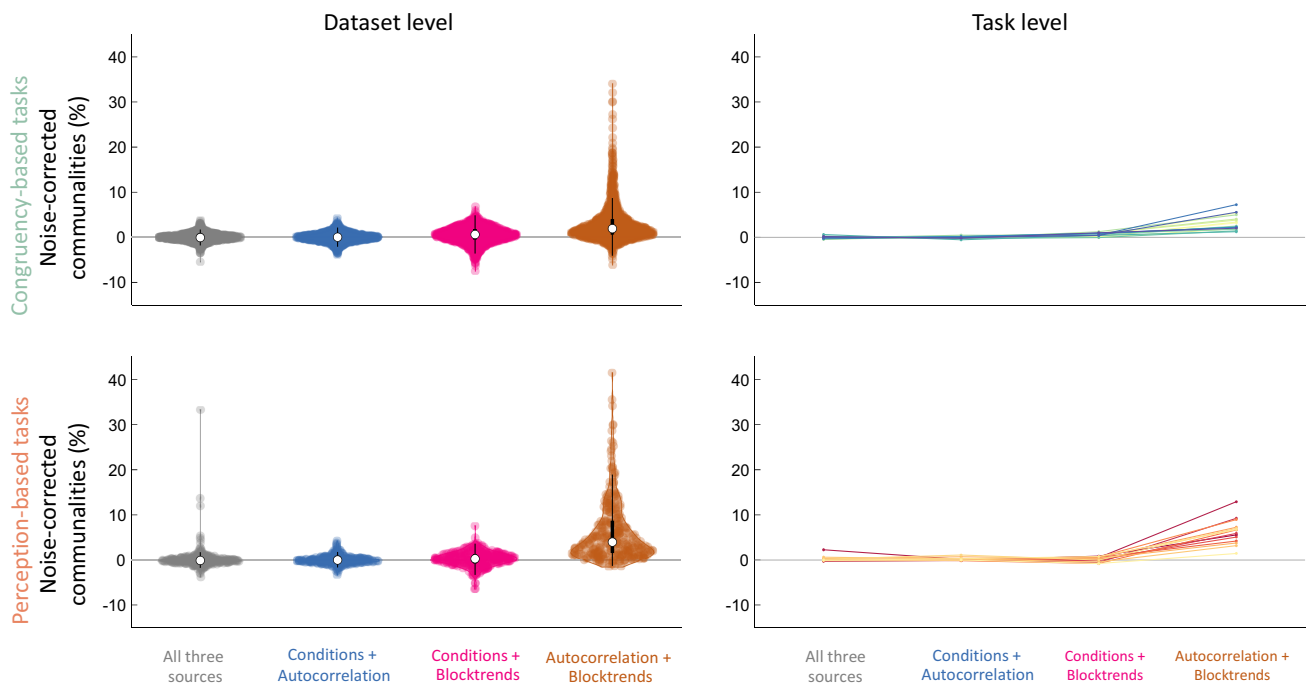
Next, as it is known that accuracy on the previous trial can influence RT, we ran an additional model that included accuracy on the previous trial (lavender violins). Compared to the full model, this increased the explained variance by 1.0% for the congruency-based and .5% for the perception-based tasks. Finally, a model including both current and previous accuracy (pink violins) explained 1.6% and 2.0% more variance compared to the full model, respectively, for congruency- and perception-based tasks compared.

Control Analyses Confirm That Results Are Robust

We conducted several further control analyses to inquire whether the above results are robust to the choices made during data analysis. First, we assessed the influence of different outlier exclusion criteria and commonly applied transformations of RT. Second, we verified that task-irrelevant exogenous elements in the design (e.g., basic stimulus characteristics or inter-trial interval durations) contribute very little to the overall RT variance. Third, the main analyses were rerun after excluding data sets for which the predictors showed signs

Figure 6

Percentage of Common Explained Variance (Communalities) by Experimental Condition, Autocorrelation, and Blockwise Trends



Note. Percentage of common explained within-subject variance by different sources: experimental conditions plus autocorrelational (blue), experimental conditions plus blockwise trends (pink), autocorrelation plus blockwise trends (brown), and all three combined (grey), shown for the congruency-based (top) and the perception-based (bottom) tasks. The communalities are shown both on the data set level (left, with each dot representing one RT series) and task level (right, with each colored line representing one task). The communalities between autocorrelation and blockwise trends are above zero, indicating an overlap of explained variance. The other three distributions are near zero, indicating negligible overlap. Communalities were calculated on the noise-corrected estimates of explained variance. RT = reaction time. See the online article for the color version of this figure.

of severe multicollinearity (marked with an asterisk in Table 2). None of these aspects affected the conclusions we have presented: Their numerical influences were small. Finally, we checked if the main results change when removing the variance of experimental conditions and blockwise trends from the autocorrelations—by first regressing RT on experimental conditions and blockwise trends, and subsequently using the residuals as predictors for the previous RTs. This procedure reduced the total explained variance by the autocorrelation, but not the unique explained variance nor the explained variance by all predictors combined—confirming the results of the communalities analysis. We report all control analyses in the Supplemental Materials C.

Mixed Models

There has been an increasing interest in linear mixed models (LMMs) to study the effects of experimental conditions (Singmann & Kellen, 2019). These can include both fixed effects (which remain constant) and random effects (which signify variance across the levels of grouping variables in the data set). In typical between-subject approaches, they can therefore explain variance between participants which GLMs will assign as error variance. When one wants to take advantage of this feature of LMMs, all participants must thus be included in one overarching model, so it can assign as much of the variability between participants as possible to the

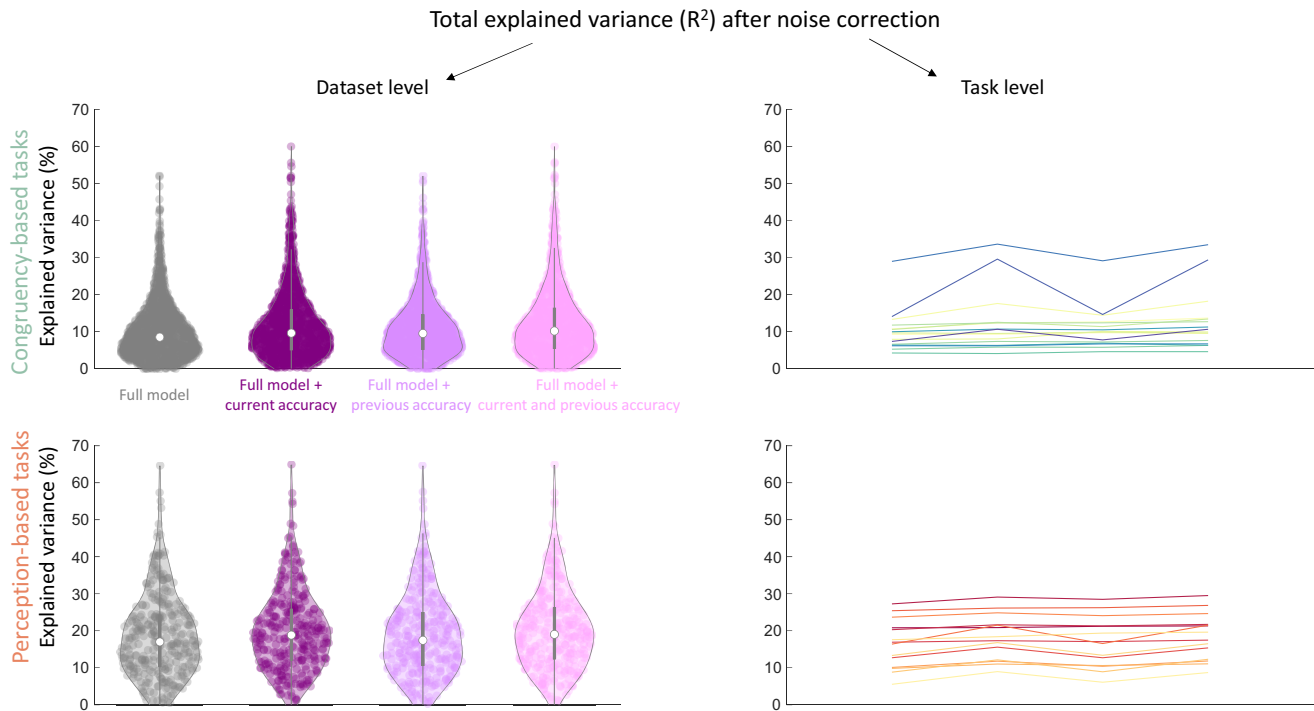
random factors. Note that in contrast, our GLM approach used subject-wise GLMs, thus fitting each participant's data independent of all others and accordingly allowing each GLM maximal flexibility to fit individual data. Therefore, employing an LMM to our multi-study/multi-participant data sets (see Figure 3) has a different focus than the analysis we have presented. We present LMM analyses for the current data in the Supplemental Materials D. The LMM explained overall more variance than the GLM approach, but still left most of the total variance unexplained. This difference was due to higher amounts of explained variance related to short- and long-term dependencies. In contrast, experimental manipulations explained less variance in the LMM than in the GLMs. We discuss potential pitfalls in these numerical comparisons between the two model approaches in section *Model form* in the Discussion.

Individual Differences in Performance Have an Unsystematic and Small Impact

The amount of explained variance showed large differences between individuals. One might speculate that these are driven by performance—for example, a well-performing participant might be less influenced by temporal effects and show larger experimental effects—or by the variability itself—for example, if a participant is more variable, there is also more potential variability to explain. To

Figure 7

Percentages of Total and Unique Explained Variance in Models Including Accuracy as Predictor



Note. Percentage of noise-corrected within-subject variance by the full model (grey, also plotted in Figure 5), by the full model plus accuracy on the current trial (dark purple), by the full model plus accuracy on the previous trial (lavender), and by the full model plus accuracy on the current and previous trial (soft pink). The explained variances are shown both on the data set level (left, with each dot representing one RT series) and task level (right, with each colored line representing one task). RT = reaction time. See the online article for the color version of this figure.

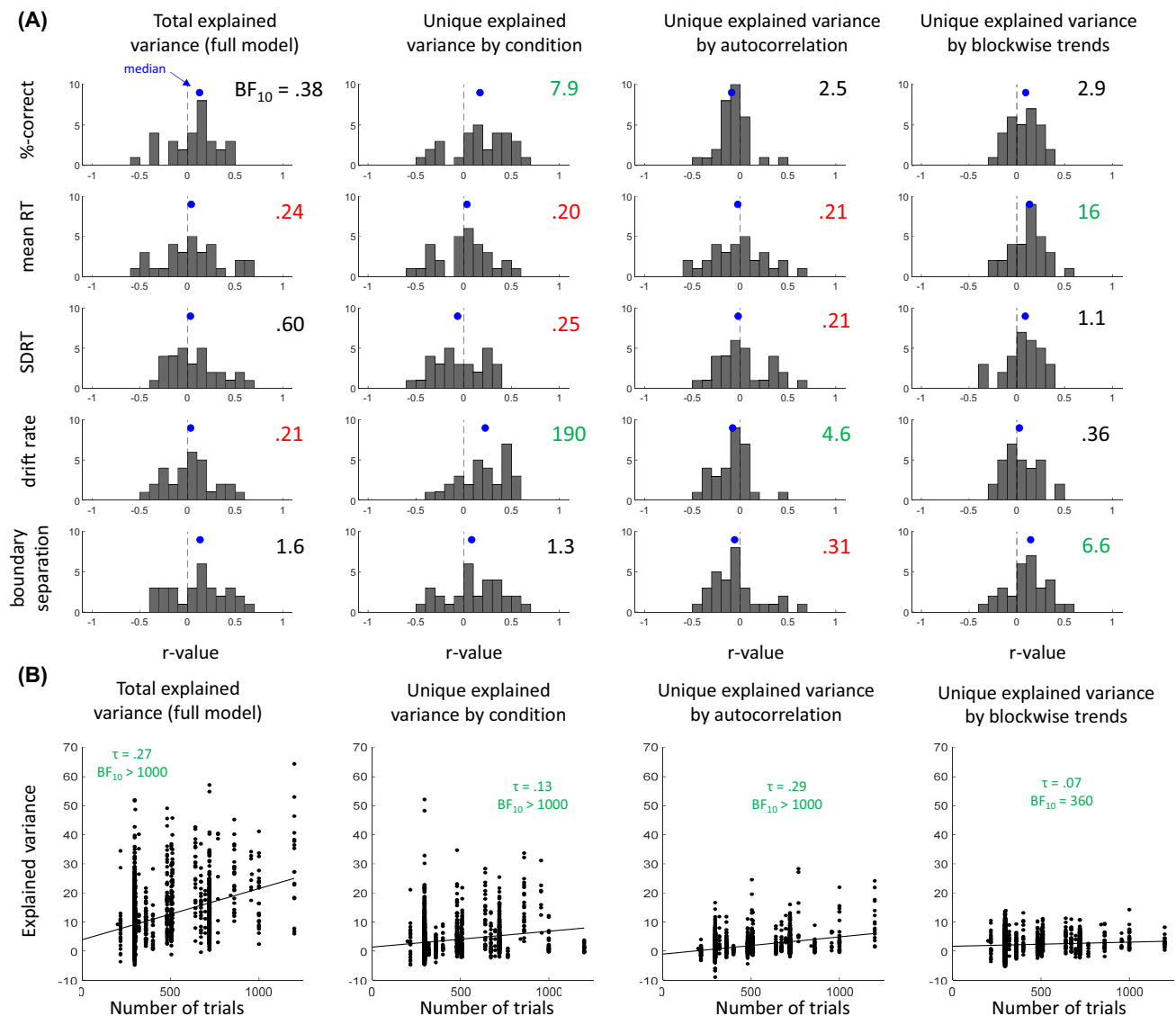
test whether individual differences in performance or speed–accuracy trade-offs might explain differences in explained variance, Pearson correlation coefficients were calculated separately per task between five measures of performance/speed–accuracy (total %-correct, mean RT on the correct trials, *SD* of RT on the correct trials, and drift rate and boundary separation as calculated by the EZ drift diffusion model) and four measures of explained variance (total by the full model, and uniquely explained by each of the three sources). If individual differences in speed–accuracy trade-offs can (partially) explain individual differences in explained variance, this would be reflected in the distributions of coefficients being different from zero.

Figure 8A shows the 20 distributions along with the Bayes Factor (BF_{10}) on top. Five of the 20 distributions showed clear Bayesian evidence in favor of a positive correlations between measures of performance and explained variance, seven distributions showed clear Bayesian evidence against correlations, and eight remained indeterminate. The largest correlation was found between drift rate and explained variance by condition (median $r = .23$). This means that on average 5.3% of the between-subject variance in explained variance by condition can be attributed to individual differences in drift rate—with higher drift rates indicating more explained variance. Explained variance for the other four distributions ranged from .6 to 2.3%. We conclude that performance might have a small impact on explained variance, though there is no clear consistency across the directions of effects.

We also tested the relationship between total trial number and explained variance, by correlating the same four measures of explained variance with total trial number across all data sets (Figure 8B). These correlations were systematically present and positive, meaning that both variance explained by experimental conditions as well as endogenous temporal dependencies were stronger with increasing trial numbers. However, the explained variance was likewise small (ranging from .5 to 7.3%).

Discussion

We systematically quantified the variance in trial-to-trial RTs explained linearly and additively by experimental conditions and short- and long-range temporal dependencies systematically across a sample of congruency- and perception-based tasks. We focused on studies that are representative of their respective fields, as they are specifically designed to study sensory and cognitive phenomena, and do not aim to increase or reduce RT variability. Specifically, GLMs were conducted as a means to test how much variance may be explained by conventional models in cognitive neuroscience. Although the experimental manipulations in these tasks typically induce robust differences between conditions on the level of averaged single subject and group data, they did not explain more variance than temporal dependencies such as autocorrelation and blockwise trends, which are traditionally considered nuisance. In addition, there was no overlap of the variance explained by

Figure 8*Between-Subject Correlations Between Explained Variance and Measures of Performance and Speed-Accuracy Trade-Off*

Note. (A) Distributions of Pearson correlation coefficients across tasks, across four measures of noise-corrected explained variance (columns) and five measures of performance and speed-accuracy trade-off. Each distribution was tested against zero (dotted line). The corresponding Bayes Factors are shown on top. Bayesian evidence favoring the presence of a correlation between explained variance and performance are shown in green ($BF_{10} > 3$), evidence favoring the absence of a correlation (null effect) are shown in red ($BF_{10} < 1/3$), and indeterminate Bayes Factors are shown in black. The median r -value of each distribution is represented by the blue dot. (B) Bayesian correlation analyses between the number trials and four measures of noise-corrected explained variance. The τ -value and BF_{10} are shown on top. RT = reaction time; BF = Bayes Factor; SDRT = standard deviation of reaction time. See the online article for the color version of this figure.

experimental conditions and temporal dependencies. Proportions of explained variance were similar across the two task sets. Hence, while RTs may change systematically over time, RT differences between conditions are stable. Perhaps most strikingly, our results show that GLMs leave the largest part of within-subject variability unexplained by the sources analyzed in the present work.

We analyzed data from two types of tasks that are often used in experimental psychology and cognitive neuroscience: congruency-based tasks and psychophysical perception tasks. Some have argued that psychophysical tasks are particularly suited for inducing robust

effects across participants (Smith & Little, 2018), given the frequent use of individually calibrated stimuli (which enhances the condition effect at the individual level) and the collection of small but well-trained samples (which reduces the error variance compared to testing naïve participants). In contrast, most congruency-based tasks employ fixed group designs and rely on larger, untrained participant samples. These differences are also reflected in the data sets analyzed here: Congruency-based tasks included more participants, while perception-based tasks featured more trials and blocks. In total, seven studies used individually calibrated stimuli, but their

variance patterns were not different from other tasks in any obvious way (see Table 3). However, we do not have the means to systematically compare the two types of tasks nor any differences between tasks in general, and only conclude that the overall conclusions hold up across both.

Experimental Effects Versus Intra-Individual Variability

Typically, significant differences between experimental conditions are interpreted in the form of “*responses are faster in condition A than in condition B.*” Such interpretations emphasize the *direction* (not the magnitude; also see J. N. Rouder & Haaf, 2021) of the difference between condition-wise *summary statistics (such as means and medians)* that are usually further *summarized across participants*. In this form, many experimental effects, including those analyzed in the present study, are considered to be highly robust even if the magnitude of a condition-difference may be small (J. N. Rouder & Haaf, 2018; J. Rouder et al., 2019). As such, these tasks are a popular means to test group distributions between conditions, though the amount of explained variance (or magnitude of effect) often remains unreported. The group distributions are typically analyzed under the assumption that the effects of interests are linear and additive (Blanca et al., 2018).

Yet our results show that even if the group-wise effect of an experimental manipulation is replicable, the moment-to-moment variability of an individual participant’s responses remains largely unpredictable with the same linear additive models. For example, a well-known cognitive congruency effect (the Stroop effect) was replicated in each of 20 independent participant pools in the cross-laboratory experiment “Many Labs 3” (Ebersole et al., 2016), and it has even been said that “Everybody Stroops” (Haaf & Rouder, 2017). If we were to test a naïve participant on this task and predicted that their average behavior will be faster in the congruent compared to the incongruent condition, our prediction would almost certainly be correct. However, as shown by the present analysis, our ability to predict the RTs of individual trials remains poor even for this particular task. Hence, the robust effects of experimental conditions in typical neurocognitive tasks do not immediately translate to explanatory power for RT at the single-trial level.

We also examined the increases in explained variance when adding accuracy of the current trial as predictor, as distributions of correct and incorrect RTs are known to differ (Bolsinova & Molenaar, 2018; Ratcliff, 1978). The increases were modest, although there were stark differences across tasks. These results should be interpreted with caution though. It is known that experimental conditions can cause changes in RT, as they can be manipulated directly by an experimenter. It is also likely that time can cause changes in RT, as their correlations cannot be explained bidirectionally (i.e., RT cannot cause changes in time). Accuracy itself is a measure of within-subject variability, so one cannot say that accuracy causes differences in RT—that would mix up the explanans and explanandum. In contrast, accuracy on the previous trial can be a possible source of within-subject variability: Participants slow down after they realize they made an incorrect response (post-error slowing; Rabbitt, 1966). The increase in explained variance by previous trial was modest. However, posterror slowing is highly task-dependent, as it requires either awareness of the error or explicit feedback. Most of the perception-based tasks specifically do not meet this requirement.

Temporal Dependency in Variability

Short- and long-range temporal dependencies explained more variance than the experimental conditions. Still their total contribution to explained variance is low, implying that knowledge of the precise temporal patterns of variability (e.g., the strength of autocorrelation) provides only poor prediction of the individual RTs.

For practical reasons, we limited the analysis on specific (first- and second-order) forms of temporal structures, though some studies have described higher-order structures in RTs. For example, one line of research has described long-range or scale-free patterns in behavior (e.g., Gilden, 2001; Kello et al., 2007; Thornton & Gilden, 2005; Van Orden et al., 2003). Often, such structure has been quantified using frequency analysis (Box et al., 2016), and some studies have claimed that long-range structures explain “a fair fraction” of the total variance, and in some cases more than the experimental conditions (Gilden, 2001, p. 54). These analysis methods come with various controversies, and the explained variance cannot be attributed to defined timescales, which limits its interpretability (see Wagenmakers et al., 2004 for a comprehensive overview; also see Farrell et al., 2006). In addition, entering such structures into an analysis of shared and unique variances is not straightforward. As such, it remains open whether higher order temporal structures explain more variance than the autocorrelations and blockwise trends investigated here.

Generalizability of Findings

For our study, we analyzed data from studies on cognitive control and sensory perception, using tasks that are well representative of their respective research fields. Proportions of explained variance were highly similar across congruency- and perception-based tasks. Whether our findings generalize to other types of tasks and disciplines remains an empirical question. However, congruency- and perception-based paradigms tend to produce condition differences in RT on the group level, and both typically include relatively high trial numbers to minimize measurement error compared to other fields (e.g., motor control, memory) and are often associated with error rates that are either low or fixed at the group level (e.g., by using staircase procedures). It seems unlikely we would find larger proportions of explained variance by experimental manipulations in tasks that do not share these characteristics. Furthermore, the potential reasons for why our behavior may be variable seem largely universal (see section *Other predictors of trial-to-trial fluctuations* below). Overall, the low proportion of explained variance may thus be a general feature of neurocognitive tasks. Similar conclusions have been drawn in animal work (Musall et al., 2019). This may reflect that the ongoing experimental task is only a small part of the plethora of activity the brain is processing at each moment in time, even in highly controlled laboratory environments.

The present analyses only include tasks that involve manual button presses. RT produced by other modalities are quantitatively different from button presses—for example, RT latencies from button presses are longer than those from saccadic eye movements. More importantly, they can be *qualitatively* different. For example, in rapid action selection tasks, manual RT had longer and more variable nondecision times—reflecting sensory processing and motor output—compared to saccadic RT (Bompas et al., 2017). If different types of responses are qualitatively different, we cannot

easily speculate if the specific proportions of explained variance found in the present study would generalize to trial-to-trial RT from other output modalities. Still, saccadic responses are also known to be highly variable (Carpenter, 1999; Noorani & Carpenter, 2016), and it may be fair to assume that large proportions of this variability would be unrelated to the task.

It is possible that temporal predictability is higher in sensorimotor actions other than the ones we scrutinized in the present study, particularly in tasks without experimental conditions such as tapping to a metronome (see Perquin, van Vugt, et al., 2023 for an overview of autocorrelations in RT series from various tasks with and without experimental conditions). However, even in these tasks without experimental conditions, the largest proportion of the RT variance remained unexplained—meaning the trial-to-trial responses were still largely unpredictable.

By combining data sets from different experiments, the present approach allows for large-scale replication; each data set can be considered as another replication of the same research question. However, not all are *independent* replications, because several data sets were acquired from the same participants who performed multiple tasks in one study (Figure 3). The potential bias of including participants multiple times is likely low because it is known that neither effects of experimental manipulations (Hedge et al., 2018; Rey-Mermet et al., 2018) nor temporal dependencies (Perquin, van Vugt, et al., 2023) generalize across tasks—that is, knowing the congruency effect or autocorrelation of one participant in a Stroop task tells us nothing about their effect or autocorrelation in the Flanker task. Still, there is an aspect of hierarchy to the included data sets that our approach does not take into account. To our knowledge, there are currently no methods that quantify variance solely at the level of a single participant while taking group-level aspects (e.g., task, duration, paradigm type) into account. Instead, we used LMMs to get a common estimate of explained variance across data sets, as well as one estimate of explained variance for each task (reported in Supplemental Materials D). With the large numbers of predictors in our present analysis, fitting LMMs is often impossible because the high number of estimated parameters and their dependencies cannot be estimated even from the huge amounts of data in our study, resulting in convergence errors during model fitting. Furthermore, the use of standardized effect sizes for LMMs is controversial, and there is no obvious equivalent to R^2 (Rights & Sterba, 2019). The LMM results should thus be taken with caution.

Explaining RT Variability

So far, we have emphasized that the analyses implemented here leave about 80%–90% of the variance unexplained. The obvious question that arises hence is: What explains the remaining variance? First, it is possible that the current sources explain large parts of RT variability but only in nonlinear structures. Second, it is possible that fluctuations in RT largely arise from fluctuations in preceding neurobiological processes not measured in the current data sets. Third, it is possible that RT variability is inherently random. These are not necessarily incompatible with each other; they might all be true to some extent. The present research does not provide empirical evidence in favor of or against any of these three explanations. We discuss all three below.

Explanation 1: We Need Better Models

One of the larger limitations of the present study is that the reported results are only as valid as our model choice. All the results and conclusions presented in the present study are based on GLMs. Such models are the bread and butter of cognitive neuroscience (Blanca et al., 2018), and our results highlight their limitations in single-trial prediction. Historically, linear models have been perceived as more interpretable, biologically plausible, and better suited at testing the types of hypotheses common in neuroscience than nonlinear models (but see Ivanova et al., 2021 for a rebuttal of this perception).

The choice of GLM makes the current results comparable across the three different sources as well as across the different data sets. However, it is certainly possible that the GLM is poor at explaining single-trial variance because it cannot capture nonlinear and nonadditive trends that might be present in the trial-to-trial RT series. Thus, our analysis approach may have missed systematic trial-level variations attributable to the experimental conditions and temporal dependencies. This is an important limitation as it is unclear how much of the possible relationship between sources and RT fluctuations are linear and how much are non-linear. Background knowledge is lacking here, as behavioral data have modeled almost exclusively with linear models. As such, there is an almost infinite number of nonlinear approaches to pick from, but a lack of a priori knowledge to select the most appropriate ones.

Approaches that may identify nonadditive effects, like generalized additive models, or nonlinear effects, like support vector machines, have seen increasing usage in neuroscience (e.g., Groot et al., 2021; Jin et al., 2019, 2020). On a practical level, such approaches are unfeasible in the present study as they commonly require regularization parameters separately for each data set that need cyclical cross-validation and re-fitting until the best fit is found. Consequently, we cannot rule out the possibility that the three sources do explain large parts of the variance in structures not captured by the GLM. If this is indeed the case, one might also wonder if nonlinear and nonadditive models may have a more important role in group analyses depending on the research goal (Ivanova et al., 2021). Exploration of nonlinear and nonadditive model approaches may explain variance that is not captured by conventional linear models and reveal nuances in relationships between variables (e.g., Bolsinova & Molenaar, 2018, showing that RT and accuracy are related to each other in nonlinear fashion).

Explanation 2: We Need to Measure Endogenous Processes

Of course, there are other potential sources of trial-to-trial fluctuations that we have not analyzed—with the most obvious being of endogenous nature. In the following we discuss two research lines that have investigated potential endogenous contributions to fluctuations in behavior.

Predicting Behavioral Fluctuations From Attentional State. One line of research explains fluctuations in performance as a result of changes in attentional state—for example, “I was slow to respond to that other car *because* I was not paying attention.” If this view were correct, attentional states would be one avenue for predicting upcoming responses (e.g., Baldwin et al., 2017). On the

one hand, attention may refer to a global state, relating to states as arousal or motivation. Indeed, when participants report to be attentionally off-task (e.g., mind wandering, drowsiness), their average RT variability in trials just before this report is higher compared to when they report to be on-task (Anderson et al., 2021; Laflamme et al., 2018; Seli et al., 2013; Thomson et al., 2014). However, effect sizes tend to be small: Fluctuations in attentional state reports can typically only explain ~2%–7% of the total trial-to-trial RT variability—though these reports appear to explain more variance in accuracy (Kane et al., 2021). Furthermore, subjective reports are typically measured retrospectively (e.g., “What was your attentional state just before this question appeared?”), but *upcoming* trial-to-trial variability as well as slow and incorrect responses seem to be consciously inaccessible (Perquin et al., 2020), meaning that metacognitive states likely have little to no predictive power.

On the other hand, attention may refer to the more specific mechanism of spatial orienting, which can be experimentally manipulated by introducing attentional cues and increasing predictability of the stimulus—though note that spatial attention can also vary from trial to trial due to changes in arousal (see Renart & Machens, 2014; Roelfsema et al., 2004). Attentional cues can decrease variability in behavior and brain activity (Arazi et al., 2019; Cohen & Maunsell, 2009, 2010; Ledberg et al., 2012). Likewise, this effect appears stronger on accuracy than on RT—for example, Azari et al. (2009) found that valid cues increased accuracy compared to invalid cues but did not significantly affect RT. Overall, while attentional state is consistently associated with behavioral variability, it is unlikely to explain large proportions of the trial-to-trial fluctuations in RT.

Predicting Behavioral Fluctuations From Neuroimaging Data. Variability is present throughout the nervous system, from phenomena such as synaptic transmission to cross-network connectivity (for comprehensive reviews, see Ermentrout et al., 2008; Faisal et al., 2008; Garrett et al., 2013; Renart & Machens, 2014; Waschke et al., 2021), and is largely independent from the task at hand (Musall et al., 2019). Variability in neural activity prior to registration of a stimulus has been associated with fluctuations in subsequent actions in response to that stimulus across human and animal studies (e.g., Cohen & Maunsell, 2009; Everling et al., 1997; Parto Dezfouli et al., 2018; Supér et al., 2003; Womelsdorf et al., 2006; Zhang et al., 2008). Correlations between fluctuations in behavior and brain state have been found across multiple time scales, from ultra-slow (Monto et al., 2008; Weissman et al., 2006) to alpha, beta, and gamma fluctuations (e.g., Bompas et al., 2015; Gonzalez Andino et al., 2005; Hamm et al., 2010; Perfetti et al., 2011). These fluctuations may influence the encoding of sensory stimuli, as evidenced by electroencephalogram (Iemi et al., 2019) and voltage imaging data (Davis et al., 2020; Liu et al., 2022). Oscillatory phase may also play a role (e.g., Busch et al., 2009; Schroeder & Lakatos, 2009; VanRullen, 2016; but for negative findings, see Benwell et al., 2017; Bompas et al., 2015; Ruzzoli et al., 2019). Functional connectivity within and between tasks, sensory, and default networks also contribute to behavioral variability (e.g., Thompson et al., 2013; Zuberer et al., 2021). Furthermore, fluctuations in behavior have been associated with other biological processes such as heartbeat (e.g., H. D. Park et al., 2014; Salomon et al., 2016), respiration (e.g., Johannknecht & Kayser, 2022), and stomach activity (e.g., Rebollo & Tallon-Baudry, 2022; Richter et al., 2017).

Collectively, this makes for a very heterogeneous body of work relying on diverse paradigms, analysis methods, and time scale. However, most studies have not reported which proportion of the total variance is explained by the respective source of interest, and endogenous and exogenous sources of variability are commonly conflated. Studies that do report proportions of explained variance point to low effect sizes: 1%–4% of endogenous saccadic RT variability could be explained by oscillatory power across the task-network (Bompas et al., 2015), while total cortical transmission times across task-networks could explain on average 8% of simple RT variability (Paraskevopoulou et al., 2021). It is thus unlikely that there is either a single process or a small number of processes that will explain a large proportion of behavioral variability.

Explanation 3: We Cannot Explain Variability

It is possible that part of moment-to-moment variability is inherently random, and thus cannot be explained from any other systematic source. By definition, such random fluctuations would be truly unpredictable.

The bar of empirical evidence for this explanation is much higher, as it would require examining all possible sources and all possible source-behavior relationships. Still, possible benefits of randomness have been highlighted previously in the literature. Random variability has been highlighted as a possible means to facilitate creative, flexible, and strategic behavior and even free will, as randomness in behavior prevents stereotypy and monotony (Carpenter, 1999; Renart & Machens, 2014). Unpredictable variability might be advantageous in contexts in which perpetuation is unwarranted, be it fleeing from a predator, exploring the environment, deciding between similar options, motor learning, or curiosity (e.g., Briffa, 2013; Chang et al., 2017; Dinstein et al., 2015; Shahan & Chase, 2002; Sternad, 2018).

Perspectives for Future Research

A comprehensive understanding of human behavior as such requires an understanding of the contributing endogenous fluctuations (see Evans et al., 2020; Uddin, 2020). Several approaches may be beneficial for addressing this question.

First, endogenous and exogenous variability are often not fully separated; any time-varying task-element (e.g., interleaved conditions, inter-trial intervals, breaks) can potentially affect behavior. When the interest lies on endogenous—that is, spontaneous—fluctuations, it is therefore important to minimize exogenous elements. This can for example be achieved by using highly repetitive tasks, such as tapping to a fixed tone or rapid action selection tasks that include easily perceivable stimuli and simple actions. When exogenous factors are still present, they will first need to be removed from the trial-to-trial fluctuations prior to data analysis (e.g., Benwell et al., 2018; Bompas et al., 2015).

Furthermore, multimodal measurements that concurrently assess neural and bodily signals will be central to identifying which of the many possible neural and bodily processes uniquely contribute to behavioral fluctuations. An important aspect for the identification of variance sources will be that studies routinely report the proportion that each source explains of the total variance on the single trial level. This requires statistical approaches that focus on single trial data rather than summary statistics, and that analyze RT as a

continuous rather than categorical (e.g., median split) variable. Such approaches may also benefit from nonlinear models (Ivanova et al., 2021).

Both behavioral variability such as the mean and variance across an entire session (Hedge et al., 2018; Hultsch et al., 2000, 2002; Perquin & Bompas, 2019; Perquin, van Vugt, et al., 2023; Saville et al., 2011, 2012), as well as temporal dependencies in entire RT series (Perquin, van Vugt, et al., 2023) exhibit considerable stability across time and are thus trait-like. This makes them well-suited for research about individual differences. Furthermore, individual differences in variance structure may be a reliable trait—that is, people reliably differ in how consistent they are on congruent versus incongruent trials in cognitive tasks (Williams et al., 2022). Various neuropsychiatric diseases have already been associated with increased variability compared to healthy populations, such as attention deficit hyperactivity disorder (Kofler et al., 2013; Tamm et al., 2012), Alzheimer's Disease (Tales et al., 2012; Tse et al., 2010), as well as schizophrenia, depression, and borderline disorder (Kaiser et al., 2008; also see Dinstein et al., 2015). Furthermore, recent work has suggested that neural activity associated with experimental factors is highly consistent across participants, while neural activity associated with fluctuations in task performance has high intersubject variability but is stable within individuals over time (Nakuci et al., 2022). These findings suggest that neural fluctuations relating to performance could in itself be an individual trait, though their reliability over time and contexts remains an open question.

More complex computational models capture variability in latent variables. For example, accumulation models (e.g., Bogacz et al., 2006; Ratcliff et al., 2016) identify a drift rate of evidence accumulation for the ongoing decision and a threshold at which decisions are committed to. These types of models have been successful at simulating RT distributions and capturing the relationship between RT and accuracy and may even be restricted with neuroimaging data (e.g., Turner et al., 2013, 2016). However, their contribution to understanding trial-to-trial fluctuations is limited, because (a) they include noise parameters agnostically fitted solely for the purpose of computational tractability, (b) they posit, by definition, that RT series are random over time, and (c) they cannot, for this reason, predict a single trial's random walk or evidence accumulation (also see Evans & Wagenmakers, 2019; Evans et al., 2020). One avenue to understanding trial-to-trial fluctuations might be to expand such models, for example, by incorporating temporal dependencies and implementing single-trial estimates of latent variables. We note that accumulation models may implement within-trial (drift rate) and between-trial (e.g., between-trial standard deviation of drift rate) variability, representing fast fluctuations in accumulation within single trials and slower fluctuations in global processes across trials respectively. However, our current analyses do not allow for a distinction of within- and between-trial variability, which would require either neural recordings (e.g., Steinemann et al., 2022) or specific experimental manipulations (e.g., Ratcliff et al., 2018; Ritz & Shenhav, 2022). Overall, the question of which convenience noise parameters relate to which neurocognitive processes remains largely open (Evans & Wagenmakers, 2019; Evans et al., 2020).

Finally, we would like to note that while our analysis indicates that the variance explained by experimental conditions and temporal dependencies are unique, this independence does not provide a rationale for refraining from the inclusion of temporal dependencies

in statistical analysis of experimental effects. Incorporating temporal dependencies can still reduce the error variance and can as such be beneficial for statistical detection of the experimental effect (e.g., Baayen & Milin, 2010; Fründ et al., 2014).

Constraints on Generality

The selected data sets in this research reflect the existing literature. As published articles in cognitive neuroscience are highly skewed toward White, Educated, Industrialized, Rich, and Democratic samples (Henrich et al., 2010), our results are likely skewed in this direction as well. The implications of this bias on the current results remain unclear as cultural differences in cognitive functions are understudied. There is evidence that visual perception can systematically differ between cultures (e.g., in the Müller-Lyer illusion; Henrich et al., 2010), so we cannot easily state that the more basic functions examined in the present work are safe from selection bias. We are not aware of any empirical evidence on the generality of temporal dependencies, but temporal dependencies seem to have mostly been estimated on White, Educated, Industrialized, Rich, and Democratic samples as well. If variability has any biological, social, or evolutionary advantages, it is not implausible that systematic difference in its structures might exist—though this remains an empirical question.

Conclusion

Experiments are a key tool for identifying factors that influence neural processing and behavior when we aggregate over trials and/or individuals, but they predict only a fraction of intra-individual variability. Endogenous sources might contribute a much higher proportion to moment-to-moment fluctuations than exogenous influences. In laboratory-derived behavior data, endogenous or unexplained variability often carries the negative connotation of simply being measurement error. Similarly, in daily life we often focus on reducing variability (e.g., in traffic). However, variability may be an intrinsic and important feature of our biological system (Faisal et al., 2008; Garrett et al., 2013). The larger empirical question is therefore to what extent trial-to-trial variability is predictable. If all possible endogenous sources of variability have been scrutinized, it is possible that some proportion of variability is truly random and thus evades precise trial-to-trial prediction.

References

- Anderson, T., Petranker, R., Lin, H., & Farb, N. A. S. (2021). The metronome response task for measuring mind wandering: Replication attempt and extension of three studies by Seli et al. *Attention, Perception, & Psychophysics*, 83(1), 315–330. <https://doi.org/10.3758/s13414-020-02131-x>
- Annand, C. T., & Holden, J. G. (2023). Embodied nonlinear dynamics of cognitive performance. *Journal of Experimental Psychology: General*, 152(5), 1264–1285. <https://doi.org/10.1037/xge0001319>
- Arazi, A., Yeshurun, Y., & Dinstein, I. (2019). Neural variability is quenched by attention. *The Journal of Neuroscience*, 39(30), 5975–5985. <https://doi.org/10.1523/JNEUROSCI.0355-19.2019>
- Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3(2), 12–28. <https://doi.org/10.21500/20112084.807>
- Baldwin, C. L., Roberts, D. M., Barragan, D., Lee, J. D., Lerner, N., & Higgins, J. S. (2017). Detecting and quantifying mind wandering during

- simulated driving. *Frontiers in Human Neuroscience*, 11, Article 406. <https://doi.org/10.3389/fnhum.2017.00406>
- Benwell, C. S. Y., Keitel, C., Harvey, M., Gross, J., & Thut, G. (2018). Trial-by-trial co-variation of pre-stimulus EEG alpha power and visuospatial bias reflects a mixture of stochastic and deterministic effects. *European Journal of Neuroscience*, 48(7), 2566–2584. <https://doi.org/10.1111/ejn.13688>
- Benwell, C. S. Y., Tagliabue, C. F., Veniero, D., Cecere, R., Savazzi, S., & Thut, G. (2017). Prestimulus EEG power predicts conscious awareness but not objective visual performance. *eNeuro*, 4(6), 1–17. <https://doi.org/10.1523/ENEURO.0182-17.2017>
- Blanca, M. J., Alarcón, R., & Bono, R. (2018). Current practices in data analysis procedures in psychology: What has changed? *Frontiers in Psychology*, 9, Article 2558. <https://doi.org/10.3389/fpsyg.2018.02558>
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, 113(4), 700–765. <https://doi.org/10.1037/0033-295X.113.4.700>
- Bolsinova, M., & Molenaar, D. (2018). Modeling nonlinear conditional dependence between response time and accuracy. *Frontiers in Psychology*, 9, Article 1525. <https://doi.org/10.3389/fpsyg.2018.01525>
- Bompas, A., Hedge, C., & Sumner, P. (2017). Speeded saccadic and manual visuo-motor decisions: Distinct processes but same principles. *Cognitive Psychology*, 94, 26–52. <https://doi.org/10.1016/j.cogpsych.2017.02.002>
- Bompas, A., Sumner, P., Muthumaraswamy, S. D., Singh, K. D., & Gilchrist, I. D. (2015). The contribution of pre-stimulus neural oscillatory activity to spontaneous response time variability. *NeuroImage*, 107, 34–45. <https://doi.org/10.1016/j.neuroimage.2014.11.057>
- Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, 73(3), 1045–1055. <https://doi.org/10.2307/1940179>
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time series analysis: Forecasting and control* (5th ed.). Wiley.
- Briffa, M. (2013). Plastic proteans: Reduced predictability in the face of predation risk in hermit crabs. *Biology Letters*, 9(5), Article 20130592. <https://doi.org/10.1098/rsbl.2013.0592>
- Busch, N. A., Dubois, J., & VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *The Journal of Neuroscience*, 29(24), 7869–7876. <https://doi.org/10.1523/JNEUROSCI.0113-09.2009>
- Carpenter, R. H. S. (1999). A neural mechanism that randomises behaviour. *Journal of Consciousness Studies*, 6(1), 13–22. <https://psycnet.apa.org/record/1999-00655-001>
- Chang, C. C., Teo, H. Y., Norma-Rashid, Y., & Li, D. (2017). Predator personality and prey behavioural predictability jointly determine foraging performance. *Scientific Reports*, 7(1), Article 40734. <https://doi.org/10.1038/srep40734>
- Chen, Y., Repp, B. H., & Patel, A. D. (2002). Spectral decomposition of variability in synchronization and continuation tapping: Comparisons between auditory and visual pacing and feedback conditions. *Human Movement Science*, 21(4), 515–532. [https://doi.org/10.1016/S0167-9457\(02\)00138-0](https://doi.org/10.1016/S0167-9457(02)00138-0)
- Cohen, M. R., & Maunsell, J. H. R. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience*, 12(12), 1594–1600. <https://doi.org/10.1038/nn.2439>
- Cohen, M. R., & Maunsell, J. H. R. (2010). A neuronal population measure of attention predicts behavioral performance on individual trials. *The Journal of Neuroscience*, 30(45), 15241–15253. <https://doi.org/10.1523/JNEUROSCI.2171-10.2010>
- Correll, J. (2008). 1/f noise and effort on implicit measures of bias. *Journal of Personality and Social Psychology*, 94(1), 48–59. <https://doi.org/10.1037/0022-3514.94.1.48>
- Davis, Z. W., Muller, L., Martinez-Trujillo, J., Sejnowski, T., & Reynolds, J. H. (2020). Spontaneous travelling cortical waves gate perception in behaving primates. *Nature*, 587(7834), 432–436. <https://doi.org/10.1038/s41586-020-2802-y>
- Desender, K., Boldt, A., Verguts, T., & Donner, T. H. (2019). Confidence predicts speed–accuracy tradeoff for subsequent decisions. *eLife*, 8, Article e43499. <https://doi.org/10.7554/eLife.43499>
- Desender, K., Van Opstal, F., Hughes, G., & Van den Bussche, E. (2016). The temporal dynamics of metacognition: Dissociating task-related activity from later metacognitive processes. *Neuropsychologia*, 82, 54–64. <https://doi.org/10.1016/j.neuropsychologia.2016.01.003>
- Desender, K., Van Opstal, F., & Van den Bussche, E. (2014). Feeling the conflict: The crucial role of conflict experience in adaptation. *Psychological Science*, 25(3), 675–683. <https://doi.org/10.1177/0956797613511468>
- Dinstein, I., Heeger, D. J., & Behrmann, M. (2015). Neural variability: Friend or foe? *Trends in Cognitive Sciences*, 19(6), 322–328. <https://doi.org/10.1016/j.tics.2015.04.005>
- Donders, A. R. T., van der Heijden, G. J. M. G., Stijnen, T., & Moons, K. G. M. (2006). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, 59(10), 1087–1091. <https://doi.org/10.1016/j.jclinepi.2006.01.014>
- Duthoo, W., Abrahamse, E. L., Braem, S., Boehler, C. N., & Notebaert, W. (2014). The heterogeneous world of congruency sequence effects: An update. *Frontiers in Psychology*, 5, Article 1001. <https://doi.org/10.3389/fpsyg.2014.01001>
- Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., Baranski, E., Bernstein, M. J., Bonfiglio, D. B. V., Boucher, L., Brown, E. R., Budiman, N. I., Cairo, A. H., Capaldi, C. A., Chartier, C. R., Chung, J. M., Cicero, D. C., Coleman, J. A., Conway, J. G., ... Nosek, B. A. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology*, 67, 68–82. <https://doi.org/10.1016/j.jesp.2015.10.012>
- Egner, T. (2017). Conflict adaptation. In T. Egner (Ed.), *The Wiley handbook of cognitive control* (pp. 64–78). Wiley. <https://doi.org/10.1002/9781118920497.ch4>
- Ermentrout, G. B., Galán, R. F., & Urban, N. N. (2008). Reliability, synchrony and noise. *Trends in Neurosciences*, 31(8), 428–434. <https://doi.org/10.1016/j.tins.2008.06.002>
- Evans, N. J., Tillman, G., & Wagenmakers, E. J. (2020). Systematic and random sources of variability in perceptual decision-making: Comment on Ratcliff, Voskuilen, and McKoon (2018). *Psychological Review*, 127(5), 932–944. <https://doi.org/10.1037/rev0000192>
- Evans, N. J., & Wagenmakers, E. J. (2019). *Evidence accumulation models: Current limitations and future directions*. PsyArXiv. <https://doi.org/10.31234/osf.io/74df9>
- Everling, S., Krappmann, P., Spantekow, A., & Flohr, H. (1997). Influence of pre-target cortical potentials on saccadic reaction times. *Experimental Brain Research*, 115(3), 479–484. <https://doi.org/10.1007/PL00005717>
- Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience*, 9(4), 292–303. <https://doi.org/10.1038/nrn2258>
- Farrell, S., Wagenmakers, E.-J., & Ratcliff, R. (2006). 1/f noise in human cognition: Is it ubiquitous, and what does it mean? *Psychonomic Bulletin & Review*, 13(4), 737–741. <https://doi.org/10.3758/BF03193989>
- Fritsche, M., Mostert, P., & de Lange, F. P. (2017). Opposite effects of recent history on perception and decision. *Current Biology*, 27(4), 590–595. <https://doi.org/10.1016/j.cub.2017.01.006>
- Fründ, I., Wichmann, F. A., & Macke, J. H. (2014). Quantifying the effect of intertrial dependence on perceptual decisions. *Journal of Vision*, 14(7), Article 9. <https://doi.org/10.1167/14.7.9>
- Garrett, D. D., Samanez-Larkin, G. R., MacDonald, S. W., Lindenberger, U., McIntosh, A. R., & Grady, C. L. (2013). Moment-to-moment brain signal variability: A next frontier in human brain mapping? *Neuroscience and Biobehavioral Reviews*, 37(4), 610–624. <https://doi.org/10.1016/j.neubio.2013.02.015>
- Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychological Review*, 108(1), 33–56. <https://doi.org/10.1037/0033-295X.108.1.33>

- Gonzalez Andino, S. L., Michel, C. M., Thut, G., Landis, T., & Grave de Peralta, R. (2005). Prediction of response speed by anticipatory high-frequency (gamma band) oscillations in the human brain. *Human Brain Mapping*, 24(1), 50–58. <https://doi.org/10.1002/hbm.20056>
- Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, 121(4), 480–506. <https://doi.org/10.1037/0096-3445.121.4.480>
- Groot, J. M., Boayue, N. M., Csifcsák, G., Boekel, W., Huster, R., Forstmann, B. U., & Mittner, M. (2021). Probing the neural signature of mind wandering with simultaneous fMRI-EEG and pupillometry. *NeuroImage*, 224, Article 117412. <https://doi.org/10.1016/j.neuroimage.2020.117412>
- Haaf, J. M., & Rouder, J. N. (2017). Developing constraint in Bayesian mixed models. *Psychological Methods*, 22(4), 779–798. <https://doi.org/10.1037/met0000156>
- Hamm, J. P., Dyckman, K. A., Ethridge, L. E., McDowell, J. E., & Clementz, B. A. (2010). Preparatory activations across a distributed cortical network determine production of express saccades in humans. *The Journal of Neuroscience*, 30(21), 7350–7357. <https://doi.org/10.1523/JNEUROSCI.0785-10.2010>
- Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 50(3), 1166–1186. <https://doi.org/10.3758/s13428-017-0935-1>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Beyond WEIRD: Towards a broad-based behavioral science. *Behavioral and Brain Sciences*, 33(2–3), 111–135. <https://doi.org/10.1017/S0140525X10000725>
- Holden, J. G., & Rajaraman, S. (2012). The self-organization of a spoken word. *Frontiers in Psychology*, 3, Article 209. <https://doi.org/10.3389/fpsyg.2012.00209>
- Hultsch, D. F., MacDonald, S. W. S., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *The Journals of Gerontology: Series B, Psychological Sciences and Social Sciences*, 57(2), 101–115. <https://doi.org/10.1093/geronb/57.2.P101>
- Hultsch, D. F., MacDonald, S. W. S., Hunter, M. A., Levy-Bencheton, J., & Strauss, E. (2000). Intraindividual variability in cognitive performance in older adults: Comparison of adults with mild dementia, adults with arthritis, and healthy adults. *Neuropsychology*, 14(4), 588–598. <https://doi.org/10.1037/0894-4105.14.4.588>
- Iemi, L., Busch, N. A., Laudini, A., Haegens, S., Samaha, J., Villringer, A., & Nikulin, V. V. (2019). Multiple mechanisms link prestimulus neural oscillations to sensory responses. *eLife*, 8, Article e43620. <https://doi.org/10.7554/eLife.43620>
- Ivanova, A. A., Schrimpf, M., Anzellotti, S., Zaslavsky, N., Fedorenko, E., & Isik, L. (2021). *Beyond linear regression: Mapping models in cognitive neuroscience should align with research goals*. bioRxiv. <https://doi.org/10.1101/2021.04.02.438248>
- JASP Team. (2017). JASP (Version 0.14.1.0).
- Jin, C. Y., Borst, J. P., & van Vugt, M. K. (2019). Predicting task-general mind-wandering with EEG. *Cognitive, Affective & Behavioral Neuroscience*, 19(4), 1059–1073. <https://doi.org/10.3758/s13415-019-00707-1>
- Jin, C. Y., Borst, J. P., & van Vugt, M. K. (2020). Distinguishing vigilance decrement and low task demands from mind-wandering: A machine learning analysis of EEG. *European Journal of Neuroscience*, 52(9), 4147–4164. <https://doi.org/10.1111/ejn.14863>
- Johannknecht, M., & Kayser, C. (2022). The influence of the respiratory cycle on reaction times in sensory-cognitive paradigms. *Scientific Reports*, 12(1), Article 2586. <https://doi.org/10.1038/s41598-022-06364-8>
- Johnson, P. C. (2014). Extension of Nakagawa & Schielzeth's R^2 GLMM to random slopes models. *Methods in Ecology and Evolution*, 5(9), 944–946. <https://doi.org/10.1111/2041-210X.12225>
- Kaiser, S., Roth, A., Rentrop, M., Friederich, H.-C., Bender, S., & Weisbrod, M. (2008). Intra-individual reaction time variability in schizophrenia, depression and borderline personality disorder. *Brain and Cognition*, 66(1), 73–82. <https://doi.org/10.1016/j.bandc.2007.05.007>
- Kane, M. J., Smeekens, B. A., Meier, M. E., Welhaf, M. S., & Phillips, N. E. (2021). Testing the construct validity of competing measurement approaches to probed mind-wandering reports. *Behavior Research Methods*, 53(6), 2372–2411. <https://doi.org/10.3758/s13428-021-01557-x>
- Karalunas, S. L., Geurts, H. M., Konrad, K., Bender, S., & Nigg, J. T. (2014). Annual research review: Reaction time variability in ADHD and autism spectrum disorders: Measurement and mechanisms of a proposed trans-diagnostic phenotype. *Journal of Child Psychology and Psychiatry*, 55(6), 685–710. <https://doi.org/10.1111/jcpp.12217>
- Karalunas, S. L., Huang-Pollock, C. L., & Nigg, J. T. (2013). Is reaction time variability in ADHD mainly at low frequencies? *Journal of Child Psychology and Psychiatry*, 54(5), 536–544. <https://doi.org/10.1111/jcpp.12028>
- Karimi-Rouzbahani, H., Woolgar, A., & Rich, A. N. (2021). Neural signatures of vigilance decrements predict behavioural errors before they occur. *eLife*, 10, Article e60563. <https://doi.org/10.7554/eLife.60563>
- Kayser, C. (2018). *Audiovisual reaction times*. Bielefeld University. <https://osf.io/pbcus/>
- Kayser, S. J., & Kayser, C. (2018). Trial by trial dependencies in multisensory perception and their correlates in dynamic brain activity. *Scientific Reports*, 8(1), Article 3742. <https://doi.org/10.1038/s41598-018-22137-8>
- Kayser, S. J., McNair, S. W., & Kayser, C. (2016). Prestimulus influences on auditory perception from sensory representations and decision processes. *Proceedings of the National Academy of Sciences of the United States of America*, 113(17), 4842–4847. <https://doi.org/10.1073/pnas.1524087113>
- Kello, C. T., Beltz, B. C., Holden, J. G., & Van Orden, G. C. (2007). The emergent coordination of cognitive function. *Journal of Experimental Psychology: General*, 136(4), 551–568. <https://doi.org/10.1037/0096-3445.136.4.551>
- Kiyonaga, A., Scimeca, J. M., Bliss, D. P., & Whitney, D. (2017). Serial dependence across perception, attention, and memory. *Trends in Cognitive Sciences*, 21(7), 493–497. <https://doi.org/10.1016/j.tics.2017.04.011>
- Kofler, M. J., Rapport, M. D., Sarver, D. E., Raiker, J. S., Orban, S. A., Friedman, L. M., & Kolomeyer, E. G. (2013). Reaction time variability in ADHD: A meta-analytic review of 319 studies. *Clinical Psychology Review*, 33(6), 795–811. <https://doi.org/10.1016/j.cpr.2013.06.001>
- Laflamme, P., Seli, P., & Smilek, D. (2018). Validating a visual version of the metronome response task. *Behavior Research Methods*, 50(4), 1503–1514. <https://doi.org/10.3758/s13428-018-1020-0>
- Ledberg, A., Montagnini, A., Coppola, R., & Bressler, S. L. (2012). Reduced variability of ongoing and evoked cortical activity leads to improved behavioral performance. *PLOS ONE*, 7(8), Article e43166. <https://doi.org/10.1371/journal.pone.0043166>
- Lemoine, L., Torre, K., & Delignières, D. (2006). Testing for the presence of 1/f noise in continuation tapping data. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale*, 60(4), 247–257. <https://doi.org/10.1037/cjep2006023>
- Liu, M., Liang, Y., Song, C., Knöpfel, T., & Zhou, C. (2022). Cortex-wide spontaneous activity non-linearly steers propagating sensory-evoked activity in awake mice. *Cell Reports*, 41(10), Article 111740. <https://doi.org/10.1016/j.celrep.2022.111740>
- Madison, G. (2004). Fractal modeling of human isochronous serial interval production. *Biological Cybernetics*, 90(2), 105–112. <https://doi.org/10.1007/s00422-003-0453-3>
- Madurski, C., & LeBel, E. P. (2015). Making sense of the noise: Replication difficulties of Correll's (2008) modulation of 1/f noise in a racial bias task. *Psychonomic Bulletin & Review*, 22(4), 1135–1141. <https://doi.org/10.3758/s13423-014-0757-4>
- Maniscalco, B., McCurdy, L. Y., Odegaard, B., & Lau, H. (2017). Limited cognitive resources explain a trade-off between perceptual and metacognitive vigilance. *The Journal of Neuroscience*, 37(5), 1213–1224. <https://doi.org/10.1523/JNEUROSCI.2271-13.2016>

- Monto, S., Palva, S., Voipio, J., & Palva, J. M. (2008). Very slow EEG fluctuations predict the dynamics of stimulus detection and oscillation amplitudes in humans. *The Journal of Neuroscience*, 28(33), 8268–8272. <https://doi.org/10.1523/JNEUROSCI.1910-08.2008>
- Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S., & Churchland, A. K. (2019). Single-trial neural dynamics are dominated by richly varied movements. *Nature Neuroscience*, 22(10), 1677–1686. <https://doi.org/10.1038/s41593-019-0502-4>
- Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society, Interface*, 14(134), Article 20170213. <https://doi.org/10.1098/rsif.2017.0213>
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- Nakuci, J., Yeon, J., Xue, K., Kim, J.-H., Kim, S.-P., & Rahnev, D. (2022). Idiosyncratic relation between human brain activity and behavior. *Biorxiv*. <https://doi.org/10.1101/2022.08.01.502338>
- Noorani, I., & Carpenter, R. H. S. (2016). The LATER model of reaction time and decision. *Neuroscience and Biobehavioral Reviews*, 64, 229–251. <https://doi.org/10.1016/j.neubiorev.2016.02.018>
- Ogundimu, E. O., Altman, D. G., & Collins, G. S. (2016). Adequate sample size for developing prediction models is not simply related to events per variable. *Journal of Clinical Epidemiology*, 76, 175–182. <https://doi.org/10.1016/j.jclinepi.2016.02.031>
- Palva, J. M., Zhigalov, A., Hirvonen, J., Korhonen, O., Linkenkaer-Hansen, K., & Palva, S. (2013). Neuronal long-range temporal correlations and avalanche dynamics are correlated with behavioral scaling laws. *Proceedings of the National Academy of Sciences of the United States of America*, 110(9), 3585–3590. <https://doi.org/10.1073/pnas.1216855110>
- Paraskevopoulou, S. E., Coon, W. G., Brunner, P., Miller, K. J., & Schalk, G. (2021). Within-subject reaction time variability: Role of cortical networks and underlying neurophysiological mechanisms. *NeuroImage*, 237, Article 118127. <https://doi.org/10.1016/j.neuroimage.2021.118127>
- Park, H., Nannt, J., & Kayser, C. (2021). Sensory- and memory-related drivers for altered ventriloquism effects and aftereffects in older adults. *Cortex*, 135, 298–310. <https://doi.org/10.1016/j.cortex.2020.12.001>
- Park, H. D., Correia, S., Ducorps, A., & Tallon-Baudry, C. (2014). Spontaneous fluctuations in neural responses to heartbeats predict visual detection. *Nature Neuroscience*, 17(4), 612–618. <https://doi.org/10.1038/nn.3671>
- Parto Dezfouli, M., Khamechian, M. B., Treue, S., Esghaei, M., & Daliri, M. R. (2018). Neural activity predicts reaction in primates long before a behavioral response. *Frontiers in Behavioral Neuroscience*, 12, Article 207. <https://doi.org/10.3389/fnbeh.2018.00207>
- Perfetti, B., Moissello, C., Landsness, E. C., Kvint, S., Pruski, A., Onofri, M., Tononi, G., & Ghilardi, M. F. (2011). Temporal evolution of oscillatory activity predicts performance in a choice-reaction time reaching task. *Journal of Neurophysiology*, 105(1), 18–27. <https://doi.org/10.1152/jn.00778.2010>
- Perquin, M. N., & Bompas, A. (2019). Reliability and correlates of intraindividual variability in the oculomotor system. *Journal of Eye Movement Research*, 12(6), Article 11. <https://doi.org/10.16910/jemr.12.6.11>
- Perquin, M. N., Heed, T., & Christoph, K. (2023, April 28). *Variance (un) explained: Experimental conditions and temporal dependencies explain similarly small proportions of reaction time variability in perceptual and cognitive tasks*. <https://osf.io/pbcus/>
- Perquin, M. N., van Vugt, M. K., Hedge, C., & Bompas, A. (2023). Temporal structure in sensorimotor variability: A stable trait, but what for? *Computational Brain & Behavior*, 6, 400–437. <https://doi.org/10.1007/s42113-022-00162-1>
- Perquin, M. N., Yang, J., Teufel, C., Sumner, P., Hedge, C., & Bompas, A. (2020). Inability to improve performance with control shows limited access to inner states. *Journal of Experimental Psychology: General*, 149(2), 249–274. <https://doi.org/10.1037/xge0000641>
- Pratte, M. S., Rouder, J. N., Morey, R. D., & Feng, C. (2010). Exploring the differences in distributional properties between Stroop and Simon effects using delta plots. *Attention, Perception, & Psychophysics*, 72(7), 2013–2025. <https://doi.org/10.3758/APP.72.7.2013>
- R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org>
- Rabbitt, P. M. (1966). Errors and error correction in choice-response tasks. *Journal of Experimental Psychology*, 71(2), 264–272. <https://doi.org/10.1037/h0022853>
- Rahnev, D., Desender, K., Lee, A. L. F., Adler, W. T., Aguilar-Lleyda, D., Akdoğan, B., Arbuzova, P., Atlas, L. Y., Balci, F., Bang, J. W., Bègue, I., Birney, D. P., Brady, T. F., Calder-Travis, J., Chetverikov, A., Clark, T. K., Davranche, K., Denison, R. N., Dildine, T. C., ... Zylberberg, A. (2020). The confidence database. *Nature Human Behaviour*, 4(3), 317–325. <https://doi.org/10.1038/s41562-019-0813-1>
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59–108. <https://doi.org/10.1037/0033-295X.85.2.59>
- Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, 20(4), 260–281. <https://doi.org/10.1016/j.tics.2016.01.007>
- Ratcliff, R., Voskuilen, C., & McKoon, G. (2018). Internal and external sources of variability in perceptual decision-making. *Psychological Review*, 125(1), 33–46. <https://doi.org/10.1037/rev0000080>
- Rebollo, I., & Tallon-Baudry, C. (2022). The sensory and motor components of the cortical hierarchy are coupled to the rhythm of the stomach during rest. *The Journal of Neuroscience*, 42(11), 2205–2220. <https://doi.org/10.1523/JNEUROSCI.1285-21.2021>
- Renart, A., & Machens, C. K. (2014). Variability in neural activity and behavior. *Current Opinion in Neurobiology*, 25, 211–220. <https://doi.org/10.1016/j.conb.2014.02.013>
- Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(4), 501–526. <https://doi.org/10.1037/xlm0000450>
- Richter, C. G., Babo-Rebelo, M., Schwartz, D., & Tallon-Baudry, C. (2017). Phase-amplitude coupling at the organism level: The amplitude of spontaneous alpha rhythm fluctuations varies with the phase of the infraslow gastric basal rhythm. *NeuroImage*, 146, 951–958. <https://doi.org/10.1016/j.neuroimage.2016.08.043>
- Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods*, 24(3), 309–338. <https://doi.org/10.1037/met0000184>
- Ritz, H., & Shenhav, A. (2022). *Humans reconfigure target and distractor processing to address distinct task demands*. *bioRxiv*. <https://doi.org/10.1101/2021.09.08.459546>
- Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (2004). Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. *Nature Neuroscience*, 7(9), 982–991. <https://doi.org/10.1038/nn1304>
- Rouder, J., Kumar, A., & Haaf, J. M. (2019). *Why most studies of individual differences with inhibition tasks are bound to fail*. *PsyArXiv*. <https://doi.org/10.31234/osf.io/3cjr5>
- Rouder, J. N., & Haaf, J. M. (2018). Power, dominance, and constraint: A note on the appeal of different design traditions. *Advances in Methods and Practices in Psychological Science*, 1(1), 19–26. <https://doi.org/10.1177/2515245917745058>
- Rouder, J. N., & Haaf, J. M. (2021). Are there reliable qualitative individual difference in cognition? *Journal of Cognition*, 4(1), Article 46. <https://doi.org/10.5334/joc.131>

- Rouder, J. N., Lu, J., Speckman, P., Sun, D., & Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin & Review*, 12(2), 195–223. <https://doi.org/10.3758/BF03257252>
- Rouder, J. N., Yue, Y., Speckman, P. L., Pratte, M. S., & Province, J. M. (2010). Gradual growth versus shape invariance in perceptual decision making. *Psychological Review*, 117(4), 1267–1274. <https://doi.org/10.1037/a0020793>
- Ruzzoli, M., Torralba, M., Morís Fernández, L., & Soto-Faraco, S. (2019). The relevance of alpha phase in human perception. *Cortex*, 120, 249–268. <https://doi.org/10.1016/j.cortex.2019.05.012>
- Salomon, R., Ronchi, R., Dönn, J., Bello-Ruiz, J., Herbelin, B., Martet, R., Faivre, N., Schaller, K., & Blanke, O. (2016). The insula mediates access to awareness of visual stimuli presented synchronously to the heartbeat. *The Journal of Neuroscience*, 36(18), 5115–5127. <https://doi.org/10.1523/JNEUROSCI.4262-15.2016>
- Samaha, J., Barrett, J. J., Sheldon, A. D., LaRocque, J. J., & Postle, B. R. (2016). Dissociating perceptual confidence from discrimination accuracy reveals no influence of metacognitive awareness on working memory. *Frontiers in Psychology*, 7, Article 851. <https://doi.org/10.3389/fpsyg.2016.00851>
- Samaha, J., & Postle, B. R. (2017). Correlated individual differences suggest a common mechanism underlying metacognition in visual perception and visual short-term memory. *Proceedings of the Royal Society B: Biological Sciences*, 284(1867), Article 20172035. <https://doi.org/10.1098/rspb.2017.2035>
- Saville, C. W. N., Pawling, R., Trullinger, M., Daley, D., Intriligator, J., & Klein, C. (2011). On the stability of instability: Optimising the reliability of intra-subject variability of reaction times. *Personality and Individual Differences*, 51(2), 148–153. <https://doi.org/10.1016/j.paid.2011.03.034>
- Saville, C. W. N., Shikhar, S., Iyengar, S., Daley, D., Intriligator, J., Boehm, S. G., Feige, B., & Klein, C. (2012). Is reaction time variability consistent across sensory modalities? Insights from latent variable analysis of single-trial P3b latencies. *Biological Psychology*, 91(2), 275–282. <https://doi.org/10.1016/j.biopsycho.2012.07.006>
- Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, 32(1), 9–18. <https://doi.org/10.1016/j.tins.2008.09.012>
- Seli, P., Cheyne, J. A., & Smilek, D. (2013). Wandering minds and wavering rhythms: Linking mind wandering and behavioral variability. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), 1–5. <https://doi.org/10.1037/a0030954>
- Shahan, T. A., & Chase, P. N. (2002). Novelty, stimulus control, and operant variability. *The Behavior Analyst*, 25(2), 175–190. <https://doi.org/10.1007/BF03392056>
- Simola, J., Zhigalov, A., Morales-Muñoz, I., Palva, J. M., & Palva, S. (2017). Critical dynamics of endogenous fluctuations predict cognitive flexibility in the Go/NoGo task. *Scientific Reports*, 7(1), Article 2909. <https://doi.org/10.1038/s41598-017-02750-9>
- Singmann, H., & Kellen, D. (2019). An introduction to mixed models for experimental psychology. D. H. Spieler & E. Schumacher (Eds.), *New methods in cognitive psychology* (pp. 4–31). Psychology Press.
- Smit, D. J. A., Linkenkaer-Hansen, K., & de Geus, E. J. C. (2013). Long-range temporal correlations in resting-state α oscillations predict human timing-error dynamics. *The Journal of Neuroscience*, 33(27), 11212–11220. <https://doi.org/10.1523/JNEUROSCI.2816-12.2013>
- Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design. *Psychonomic Bulletin & Review*, 25(6), 2083–2101. <https://doi.org/10.3758/s13423-018-1451-8>
- Steinemann, N. A., Stine, G. M., Trautmann, E. M., Zylberberg, A., Wolpert, D. M., & Shadlen, M. N. (2022). *Direct observation of the neural computations underlying a single decision*. bioRxiv. <https://doi.org/10.1101/2022.05.02.490321>
- Sternad, D. (2018). It's not (only) the mean that matters: Variability, noise and exploration in skill learning. *Current Opinion in Behavioral Sciences*, 20, 183–195. <https://doi.org/10.1016/j.cobeha.2018.01.004>
- Supér, H., van der Togt, C., Spekreijse, H., & Lamme, V. A. (2003). Internal state of monkey primary visual cortex (V1) predicts figure-ground perception. *The Journal of Neuroscience*, 23(8), 3407–3414. <https://doi.org/10.1523/JNEUROSCI.23-08-03407.2003>
- Tales, A., Leonards, U., Bompas, A., Snowden, R. J., Philips, M., Porter, G., Haworth, J., Wilcock, G., & Bayer, A. (2012). Intra-individual reaction time variability in amnesic mild cognitive impairment: A precursor to dementia? *Journal of Alzheimer's Disease*, 32(2), 457–466. <https://doi.org/10.3233/JAD-2012-120505>
- Tamm, L., Narad, M. E., Antonini, T. N., O'Brien, K. M., Hawk, L. W., Jr., & Epstein, J. N. (2012). Reaction time variability in ADHD: A review. *Neurotherapeutics*, 9(3), 500–508. <https://doi.org/10.1007/s13311-012-0138-5>
- Temple, J. G., Warm, J. S., Dember, W. N., Jones, K. S., LaGrange, C. M., & Matthews, G. (2000). The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task. *Human Factors*, 42(2), 183–194. <https://doi.org/10.1518/001872000779656480>
- Teuchies, M., Desender, K., De Baene, W., Demanet, J., & Brass, M. (2019). *Metacognitive awareness of difficulty in action selection: The role of the cingulo-opercular network*. bioRxiv. <https://doi.org/10.1101/641340>
- The MathWorks. (2021). *Release 2021b*. <https://www.mathworks.com/company/newsroom/mathworks-introduces-release-2021b-of-matlab-and-simulink.html>
- Thompson, G. J., Magnuson, M. E., Merritt, M. D., Schwarb, H., Pan, W. J., McKinley, A., Tripp, L. D., Schumacher, E. H., & Keilholz, S. D. (2013). Short-time windows of correlation between large-scale functional brain networks predict vigilance intraindividually and interindividually. *Human Brain Mapping*, 34(12), 3280–3298. <https://doi.org/10.1002/hbm.22140>
- Thomson, D. R., Seli, P., Besner, D., & Smilek, D. (2014). On the link between mind wandering and task performance over time. *Consciousness and Cognition*, 27, 14–26. <https://doi.org/10.1016/j.concog.2014.04.001>
- Thornton, T. L., & Gilden, D. L. (2005). Provenance of correlations in psychological data. *Psychonomic Bulletin & Review*, 12(3), 409–441. <https://doi.org/10.3758/BF03193785>
- Torre, K., Vergotte, G., Viel, É., Perrey, S., & Dupeyron, A. (2019). Fractal properties in sensorimotor variability unveil internal adaptations of the organism before symptomatic functional decline. *Scientific Reports*, 9(1), Article 15736. <https://doi.org/10.1038/s41598-019-52091-y>
- Torre, K., & Wagenmakers, E.-J. (2009). Theories and models for 1/f(β) noise in human movement science. *Human Movement Science*, 28(3), 297–318. <https://doi.org/10.1016/j.humov.2009.01.001>
- Tse, C.-S., Balota, D. A., Yap, M. J., Duchek, J. M., & McCabe, D. P. (2010). Effects of healthy aging and early stage dementia of the Alzheimer's type on components of response time distributions in three attention tasks. *Neuropsychology*, 24(3), 300–315. <https://doi.org/10.1037/a0018274>
- Turner, B. M., Forstmann, B. U., Wagenmakers, E. J., Brown, S. D., Sederberg, P. B., & Steyvers, M. (2013). A Bayesian framework for simultaneously modeling neural and behavioral data. *NeuroImage*, 72, 193–206. <https://doi.org/10.1016/j.neuroimage.2013.01.048>
- Turner, B. M., Rodriguez, C. A., Norcia, T. M., McClure, S. M., & Steyvers, M. (2016). Why more is better: Simultaneous modeling of EEG, fMRI, and behavioral data. *NeuroImage*, 128, 96–115. <https://doi.org/10.1016/j.neuroimage.2015.12.030>
- Uddin, L. Q. (2020). Bring the noise: Reconceptualizing spontaneous neural activity. *Trends in Cognitive Sciences*, 24(9), 734–746. <https://doi.org/10.1016/j.tics.2020.06.003>
- Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, 132(3), 331–350. <https://doi.org/10.1037/0096-3445.132.3.331>

- van Ravenzwaaij, D., Donkin, C., & Vandekerckhove, J. (2017). The EZ diffusion model provides a powerful test of simple empirical effects. *Psychonomic Bulletin & Review*, 24(2), 547–556. <https://doi.org/10.3758/s13423-016-1081-y>
- van Ravenzwaaij, D., & Oberauer, K. (2009). How to use the diffusion model: Parameter recovery of three methods: EZ, fast-dm, and DMAT. *Journal of Mathematical Psychology*, 53(6), 463–473. <https://doi.org/10.1016/j.jmp.2009.09.004>
- VanRullen, R. (2016). Perceptual cycles. *Trends in Cognitive Sciences*, 20(10), 723–735. <https://doi.org/10.1016/j.tics.2016.07.006>
- Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation and interpretation of 1/f noise in human cognition. *Psychonomic Bulletin & Review*, 11(4), 579–615. <https://doi.org/10.3758/BF03196615>
- Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, 14(1), 3–22. <https://doi.org/10.3758/BF03194023>
- Waschke, L., Kloosterman, N. A., Obleser, J., & Garrett, D. D. (2021). Behavior needs neural variability. *Neuron*, 109(5), 751–766. <https://doi.org/10.1016/j.neuron.2021.01.023>
- Weissman, D. H., Roberts, K. C., Visscher, K. M., & Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nature Neuroscience*, 9(7), 971–978. <https://doi.org/10.1038/nn1727>
- Welford, A. T. (1968). *Fundamentals of skill*. Methuen Publishing.
- Williams, D. R., Martin, S. R., & Rast, P. (2022). Putting the individual into reliability: Bayesian testing of homogeneous within-person variance in hierarchical models. *Behaviour Research Methods*, 54, 1272–1290. <https://doi.org/10.3758/s13428-021-01646-x>
- Womelsdorf, T., Fries, P., Mitra, P. P., & Desimone, R. (2006). Gamma-band synchronization in visual cortex predicts speed of change detection. *Nature*, 439(7077), 733–736. <https://doi.org/10.1038/nature04258>
- Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12(6), 1100–1122. <https://doi.org/10.1177/1745691617693393>
- Zhang, Y., Wang, X., Bressler, S. L., Chen, Y., & Ding, M. (2008). Prestimulus cortical activity is correlated with speed of visuomotor processing. *Journal of Cognitive Neuroscience*, 20(10), 1915–1925. <https://doi.org/10.1162/jocn.2008.20132>
- Zuberer, A., Kucyi, A., Yamashita, A., Wu, C. M., Walter, M., Valera, E. M., & Esterman, M. (2021). Integration and segregation across large-scale intrinsic brain networks as a marker of sustained attention and task-unrelated thought. *NeuroImage*, 229, Article 117610. <https://doi.org/10.1016/j.neuroimage.2020.117610>

Received February 7, 2023

Revision received April 18, 2024

Accepted June 4, 2024 ■