

Interaction-Dominant Causation in Mind and Brain, and Its Implication for Questions of Generalization and Replication

Sebastian Wallot¹  · Damian G. Kelty-Stephen²

Received: 28 February 2017 / Accepted: 23 November 2017 / Published online: 29 November 2017
© The Author(s) 2017. This article is an open access publication

Abstract The dominant assumption about the causal architecture of the mind is, that it is composed of a stable set of components that contribute independently to relevant observables that are employed to measure cognitive activity. This view has been called component-dominant dynamics. An alternative has been proposed, according to which the different components are not independent, but fundamentally interdependent, and are not stable basic properties of the mind, but rather an emergent feature of the mind given a particular task context. This view has been called interaction-dominant dynamics. In this paper, we review evidence for interaction-dominant dynamics as the causal architecture of the mind. We point out, that such an architecture is consistent with problems of convergence in research on the level of results and theorizing. Moreover, we point out that if interaction-dominant dynamics as the causal architecture of the mind were to be true, this would naturally lead to (some degree of) problems with generalization and replicability in sciences of the mind and brain, and would probably warrant changes in the scientific practice with regard to study-design and data analysis.

Keywords Component-dominant dynamics · Interaction-dominant dynamics · Multi-fractality · Generalization · Replication

✉ Sebastian Wallot
Sebastian.wallot@aesthetics.mpg.de

¹ Max Planck Institute for Empirical Aesthetics, Grüneburgweg 14, 60322 Frankfurt, Germany

² Department of Psychology, Grinnell College, Grinnell, IA, USA

1 Introduction

The replication crisis in psychology has prompted psychologists to engage in a great deal of soul-searching and improvement of scientific practice. We support all movement to amend sub-optimal practices such as “*p*-hacking” and to bring to light all experimenter degrees of freedom preventing principled theory-driven research (Simmons et al. 2011). However, we find it noteworthy that all of this soul-searching has largely remained in the realm of methodological cleanliness *within* the current paradigm, specifically with regard to its assumptions about measurement and analysis. It has broached less often the question of what theoretical premises prompt the original expectation that—under optimal methodological conditions—measured behaviors can be replicated at all? Biological, chemical, and physical sciences alike have had to grapple with the empirical quandary that even so-called simple one-dimensional recursive systems can become, with nothing but a linear increase in a single parameter, mathematically unpredictable (Hofstadter 1981). Psychologists appear to expect replicability out of the complex biological systems pressing the response buttons in our experiments. Looking strictly at the high dimensionality of the human, cognitive systems and contrasting that with the low-dimensionality of potentially unpredictable systems, we offer the possibility that this expectation of replicability might benefit from some narrower bounds.

Indeed, some of our cognitive psychology research might be most relevant to understanding our replicability crisis. We find a particular sentence in this article illuminating: “When the optimal strategy in a task is to provide a series of independent and identically distributed responses, people often perform sub-optimally” (Brown and Steyvers 2009, p. 50). Let us also restate our hope that scientists should all aspire to ethical and honest practices, but all that said, the optimal strategy in standard psychological methodologies has been to measure as series of responses as if they were independent and identically distributed, and if we know people do sub-optimally at such a task, replication should perhaps not be among the first expectations, as its premise rests on this very assumption—and assumption which need not hold.

Let us be clear here: we do not mean to set fire to the scientific house and cast replicability to the wind. Linear statistics are useful despite their strange assumptions of independence and identical distributions—also because of the lack of alternatives, as no other statistical toolbox has been composed that allows to summarize and evaluate empirical results similarly well. However, we do mean to clarify that psychology may be hoping for replications in the very measurements and under the very conditions where a growing body of psychological evidence explicitly discourages predicting replication. Failures to replicate may be plainly baked into the potentially incomplete, but broadly sweeping failure of human behavior to conform to the standard of independent and identically distributed behavior. We can call that feature a failure, or we might more positively recognize it as the very context-sensitivity that we regularly credit to the human species’ exceptional intelligence (Pinker 1997).

In what follows, we want to discuss two views on measurement and analysis in the sciences that investigate human behavior and neurophysiology, and their implications for how causality is viewed. According to Van Orden et al. (2003), the majority of research in those sciences subscribes to the view of causation as component dominant dynamics (CDD), where the dynamics of behavioral and neurophysiological measurements are “dominated” by a finite set of time-invariant components to which observed variance can be attributed. For example, behavioral and neurophysiological measurements during reading of this paper can in parts be attributed to distinct mental components (long-term memory, working memory, perceptual memory...) and neurophysiological components (distinct brain regions, networks).

Hence, explaining observed dynamics in brain and behavior means to identify and chart the different components involved. The notion of causation implied here is that mind and brain work in a way so that some input for a participant in an experiment is always mediated by the same components in the same (qualitative) manner to link this input to some measured output—simple feedback loops notwithstanding. The causal chain is hard-assembled.

An alternative view of causation is interaction-dominant dynamics (IDD), where component-causal chains are not seen as the causal building blocks of mind and brain dynamics, but are themselves soft-assembled. Softassembly here means that the causal components themselves emerge in the context of a particular task environment in which organism acts, and do not exist independently of that activity. As the reader might already be able to guess here, the former view (CDD) invites stronger assumptions about and demands on the stability—and hence replicability—of observed behavior than the latter (IDD).

Crucially, these two perspectives on causation—as well as the expectations they invite about the stability of observed effects—hinge on properties of measurement and assumptions about data analysis. The current mainstream view hinges on the assumption of component dominant dynamics (CDD). One of the key-measurement-assumptions for CDD is the above stated assumption that consecutively measured values of a behavioral or physiological process are independent and identically distributed (iid). The iid-assumption is the premise on which variance of a measured observable can be parsed into distinct sources of variance, which in turn can be interpreted as cognitive or neurophysiological components—the component causes of the observed dynamics.

The alternative view, interaction-dominant dynamics (IDD), relaxes the iid-assumption. Briefly summarized, is the assumption that consecutively measured values of a behavioral or physiological process are *interdependent* (i.e., violate the iid-assumption) in a way so that variance of a measured observable *cannot* be parsed into distinct sources of variance, and hence an identification of causal components that are the specific, stable, generalizable sources of such variance is not permissible—or only permissible within certain, strict boundary conditions.

So the overarching question of this paper is, how do basic issues of measurement and analysis relate to the way that explanations are formulated given a certain model of causality, and what are the implications of this relationship for questions of generalization and replicability.

In the following sections we will introduce the concepts of component- and interaction-dominant dynamics as two concepts of causation, and highlight its dependence on the statistical models that are used to analyze data in psychological and neurophysiological research. We will describe two three types of evidence for interaction-dominant causation found in behavioral and neurophysiological data. Finally, we will discuss the implications of interaction-dominant dynamics for generalization and replication of research results, which is that replication and generalization rely on stability assumptions about observed mental components, and in turn about the statistical nature of observables which are thought to reflect the activity of such components. If human behavior indeed thrives on interaction-dominant dynamics, then such stability assumptions might often not be met.

2 Component-Dominant Dynamics

The concept of component-dominant dynamics (Van Orden and Paap 1997; Van Orden et al. 2003) summarizes the default assumptions about the causal architecture of the mind/brain from an epistemological stand point. Practically, a researcher who investigates human behavior is confronted with the following situation: She possesses a data record of an observable that was measured during a particular task or activity the organism performed. In order to draw conclusions from the data, the data need to be reduced. The reduction needs to be performed in a way that— together with prior knowledge or assumptions about the observable, the organism, and the task—one can say something about what the organism actually did and how. With reduction, we specifically mean statistical analysis of the data, on which a communicable interpretation is based.

Clearly, data analysis (and recording) are itself laden with assumptions or auxiliary theories about the phenomenon of interest (Feyerabend 2010/1975), and the concept of component-dominant dynamics summarizes the following assumptions that are made. The assumptions pertaining to the explanation of the data is, that the variation in the measured record of an observable (i.e., “dynamics”) can be reduced to (i.e., are “dominated” by) a set of independent and time-invariant sources (“components”). The components are usually thought to be (neuro-)psychological constructs that describe cognitive functions, and these functions explain the variation that is observed in a measured record. The (hypothesized) nature of these constructs allows to formulate hypotheses and make predictions about to-be observed empirical results. Note that in the current mainstream, constructs are hard-assembled entities. That is, they are assumed to be a fundamental, non-changing part of the mental architecture, and contribute their effects on observed behavior in the same way across different situations. Hence, a data analysis strategy needs to be pursued that realizes a partitioning of variation, and then an explanation is sought which provides appropriate labels for the resulting component sources of variation that are interpreted in terms of psychological constructs (i.e., “memory”, “perception”, “motivation”...).

The statistical methods used in the behavioral and brain sciences—the inferential statistics that generate the picture of the results that is interpreted—are (part of) the

general linear model: Analysis of variance, correlation/regression, or more modern variants that combine the two in various ways. What they have in common is that they take an independent factor or predictor variable that is used to partition out a portion of the variability in an observable, and attribute that variability to this independent factor or predictor. And this independent variable is interpreted as reflecting the causal-component that is the source of the observed variability—i.e., the effect that the participating explanatory construct has on the measured data.

This leads us to the second set of assumptions that are summarized in component-dominant dynamics, which relate to measurement and analysis. As Van Orden et al. (2003) noted, however, analysis of the general linear model-type will always attempt to partition the variability to independent sources and hence will force an interpretation of the variability as one that is the result of multiple, independent component-sources, no matter whether such independent components exist in the first place. It is simply what the analysis does, and if we observe that such an analysis has successfully attributed a meaningful portion of the variability in an observable to a specific factor, this does not mean that his factors exists as such—rather, the analysis assumes that the structure of the data is one of multiple independent component sources.¹ Particularly, this iid-assumption (i.e., independent and identically distributed data) needs to hold for a given data record: Before general-model type of statistics can be employed to meaningfully parse a data recorded into variance components, the structure of the data need to conform to this assumption. Otherwise, partitioning a data record into distinct sources of variation is not permitted, and hence interpretation of variation in that record being caused by specific component sources it not warranted.

However, there is another measurement-related assumption summarized in CDD that considers not only demands on data within a single record or study, but also the stability of results across multiple studies. In addition, there is another assumption that is important from a practical point of view, which is that the number of components that make-up the causal architecture is a stable, finite, and the number is “reasonably limited”. Otherwise, if the number of components is changing across studies and is extremely large, then it becomes difficult to formulate a construct in a way that one can systematically draw predictions from them, and hence use them as explanations for empirical data. If the set and/or nature of the constructs need to be adjusted too quickly across multiple studies, then the scientific explanation they aim at delivering begin to rival the complexity of the un-reduced raw data that was collected to begin with.

The last-mentioned assumptions are, where replicability (and generalization) enter the scene. Let us say a researcher is interested in investigating a psychological construct, such as working memory. The construct is latent (i.e., not directly observable), but the way it is described (i.e., which functions working memory has, and how performs those functions) leads to hypotheses that predict outcomes one can observe in a data record given circumstance (i.e., an experimental setting). Now let us say, the researcher has gathered some data that yields evidence in favor of the

¹ Note that this problem also extends to the design of studies, especially experiments testing different, discrete conditions, but in this article we will focus on the analysis part only.

hypothesis, and this in favor of the construct in its particular formulation. How can things continue from here?

First, we could have perfect replicability, so performing the same experiment again yield the exact same results, not matter of often this is done. Then, the construct seems to provide a reliable explanation for memory performance. This is also the basis for generalization, because only when results are replicable can they also meaningfully generalize to other situations (i.e., other experimental setups, participants, measures etc....). A second alternative could be a moderate form of progress, where further studies lead to an incremental refinement of the construct based on empirical evidence, which converts to a stable solution, or to the formulation of a completely new construct, which can explain everything the old construct could explain, in addition to new, unaccounted observations, and which is only marginally more complex than the original construct. A third alternative, however, be non-replicability and/or convergence of results, where previous results cannot be replicated (i.e., the explanation provided by the construct is not reliable) or where they cannot be—or only sometimes be—obtained over other variations of the experiment (i.e., the explanation provided by the construct is limited and/or its reach is unclear).²

In the latter case, there are a couple of possibilities for how to proceed. One way could be that the construct (or its formulation) is simply wrong, and we need to seek a new explanation that—hopefully—does not run into the same problems. Or we need to find more appropriate experimental manipulations and measurements that capture the underlying construct better [it is actually difficult to distinguish these two situations, and we will come back to this later when outlining additive factor logic (Sternberg 1969)]. Another way could be to question the scientific practice within the current paradigm, for example faulty use of statistics, publication biases, partial reports of data etc. However, a last alternative could be that the underlying architecture of the cognitive systems simply does not conform to CDD, that is that there are no stable, finite sets of independent constructs encapsulated in the head that explains the observed data.

Sticking with the options to look for new/other constructs, “better” experimental manipulations and measurements, or improvements of the quality of the scientific practice means to accept the mainstream view of CDD, and trying to fix its problems within this paradigm. If one in in principle ready to reject the CDD view, and wants to seek for information as to whether this view and its basic assumptions are justified, what evidence could be sought out? After all, the constructs—and their whole architecture—is latent, and defies direct investigation and, as described above, failure of hypotheses, replication and generalization can be addressed wholly within the current paradigm.

² Note, however, the replicability and generalization are somewhat more intertwined in the area of psychological research, because the exact same circumstances can never be reproduced twice—at the very minimum, one deals with the same participants, but who now have had prior experience. Or with different participants, who—for examples as being part of a cohort—might bring different predispositions to the table. However, especially over longer stretches of time, a certain variation in the experimental setup or measurement apparatus (due to technological development or generally availability of certain resources) is almost inevitable.

However, such evidence can be sought on the shallow and deeper measurement-related assumptions of CDD, and that is whether there is convergence of results across studies through pervasive statistical interaction effects (Abraham 1987), or strong violations of the iid-assumption in though complexity characteristics (Jensen 1998) in measures of brain and behavior. In the former case, one does simply never obtain a stable set of results across studies, no matter the change in theory and observables. In the latter case, the statistical characteristics of the measured observables themselves principally defy partitioning into component sources. Hence, such a partitioning of variation cannot be conducted, and underlying constructs that are the source of such variation cannot be proposed.

3 Interaction Effects in the Linear Model

The non-convergence of results across studies—which is also part of the symptomology of problems with replication and generalization—can be understood in terms of statistical interaction effect. Statistical interaction effects indicate that some factor that was investigated in, say, two experiments A and B does not play the same role in A than it did in B. For sciences that concern themselves with human behavior, the work of Franciscus Donders (1869) on mental chronometry can be taken as a methodological starting point here.

Donders' worked on a procedure that would allow for an experimental distinction of different mental faculties: Using a reaction time task, he found that simple reaction times are shorter than recognition reaction times, which in turn are shorter than choice reaction time. Using a subtraction procedure, this lead him to suggest that the different tasks involved different mental faculties which are charted on the same dimension (reaction time), and that subtracting the different reaction times observed for those tasks allowed for an identification of the different faculties and their role in the respective tasks. Soon, this subtraction method would be enhanced by Fisherian statistics, which allowed for more rigorous testing of complex experimental designs, and also clearly stated the requirement that the underlying data needed to be the results of independent and identically distributed sources for such tests to be effective and interpretable (the iid-assumption).

Herbert Simon (1973) formulated the deep requirements on the part of the mental architecture, namely that the different levels of the physiological-cognitive system need to be vertically (i.e., across different time-scales) near-decomposable in order to be horizontally (i.e., on the same time-scale) decomposable. In other words, mental processes that are investigated in this manner need to play-out on a “certain” time-scale, and that different time-scales are separable (note that this does not touch on the issue of sequential or parallel processing—also parallel processing can satisfy this requirement, if the parallel processes are independent of each other and each unfold on a certain time-scale). Building on Simon's (1973) requirements, and combining Donders' (1869) suggestion with the requirements of Fisherian statistics, Saul Sternberg (1969) provided a clear, systematic logic of how analysis of behavioral data should proceed and be interpreted accordingly. Particularly, Sternberg's paper explicates the above-mentioned interdependence between the

assumption that independent components making up the architecture of mind and behavior, and experimental manipulations that target different components, as well as analysis of the same.

Suppose one wants to find the relevant cognitive components C for a particular task T . Let us suppose the task is comprised of three components, $C1$, $C2$ and $C3$. According to Sternberg's logic, we can find those components through the variation of experimental factors F , that target specifically and exclusively on of these components. For example, factor $F1$ targets specifically component $C1$, $F2$ targets $C2$, and $F3$ targets $C3$. However, there are also factors that do not target one of the components at all (they are uninformative), or there are factors that target more than one component (which suggests that more/other factors need to be investigated in order to get a clear picture of the different components, and/or suggests that there are more components than one has thought initially). For example, factor $F4$ targets components $C2$ and $C3$ simultaneously.

Minimally, we need to investigate two factors with two levels each to dissociate different components. For example, the factor combination $F1^+$ and $F1^-$, with $F2^+$ and $F2^-$ might yield the following result on an observable O : Moving from $F1^-$ to $F1^+$ increases values of O independently of the level of $F2$. Conversely, Moving from $F2^-$ to $F2^+$ increases values of O independently of the level of $F2$. Two components, $C1$ and $C2$ are dissociated. What if we want to investigate further factors, and introduce variable $F4$? Now we see, that variations in $F2$ are not independent of variations in $F4$. Moving from $F4^-$ to $F4^+$ will also change values of $F2$. This is called an interaction effect: The two factors are not independent of each other, but interact. This could imply the following courses the research could take:

1. We search for a new factor that will target $C3$ specifically, and find factor $F3$ which does so. This leads to a revision of our explanation of the behavior during task T , moving from an explanation of $T = C1 + C2$ to $T = C1 + C2 + C3$.
2. We search of a new factor that will target $C3$ specifically, but fail to do so. Then we have to revise our theory to include special cases, where the task T works differently under different conditions: $T_{F4}^- = C1 + C2_{F4}^-$, $T_{F4}^+ = C1 + C2_{F4}^+$.

In both cases, encountering an interaction effect leads to a complexification of our explanation for task T . If we investigate further factors that lead to further interaction effects, we will need to complexify our explanation even further—either through addition of components, or through exceptions, and our explanation of the data will never settle to a finite set of constructs.

The worry of non-convergence of findings in the behavioral and brain sciences has been voiced from different perspectives at different times: For example, Sullivan (2016a, b) recently described problems of non-convergent research anchored on the notion of construct stabilization. As described above, constructs are primary explanatory vehicles (theories, hypotheses, phenomenological description of mental faculties, auxiliary—often measurement related—aspects)—usually formulated in the component-sense outlined above. Ideally, researchers within a discipline, or across disciplines that work on the similar questions (i.e., psychology

and neuroscience) agree on all of these levels: What defined the phenomenon, what are relevant theories to explain the phenomenon, which procedures are suited for empirical research on the phenomenon?

Hence, stable constructs are one of the mandatory footholds for concentrated research efforts, but also simply for sensibly relating different findings across different studies and fields. However, as Sullivan concludes for psychology and neuroscience, requirements for construct stability are often not met, and research in these fields often has not stable referents and experimental practices. Quoting Poldrack (2010), she (Sullivan 2016a) summarized that “(...) the task comparisons in many studies are based on intuitive judgements regarding the cognitive processes engaged by a particular task” (p. 149)—not by a coherent research framework.

But what is the reason for this diversity in doing research? Why do different research groups not agree on the defining aspects of phenomenon, theory, and operationalization? Van Orden, Pennington, and Stone (2001—see also Van Orden and Kloos 2003) attempt to give an answer related to the statistical aspects we discussed above, and the answer can be summarized as follows: Given the current practice of data gathering and data analysis, everybody has good reasons to stick to their own, idiosyncratic paradigms, definitions, and choice of the “most adequate” measurement. In particular, they describe the development of constructions in reading research (especially related to dual-route theory) to make their case. Dual-route theory of reading provided a solution to a long-standing debate of whether reading proceeded by recognition of individual letters, or by reading each word holistically. In a nutshell, dual-route theory integrated both explanations, proposing that both routes are possible: Fast lexical-holistic word recognition, and slow, nonlexical-letter-based word recognition. Further evidence for these two routes, came from research on dyslexic patients, some of whom (called deep dyslexics) could not name pseudowords (BINT), but exceptions (PINT), while others (called surface dyslexics) could name pseudowords (BINT), but produced regularization errors to exceptions (PINT) (Marshall and Newcombe 1973). This seemed to suggest a strong neural underpinning in the form of two neural components for the two routes to word recognition.

We do not want to review the whole article here. The point is, as Van Orden et al. (2001) continued, that the distinction between deep dyslexia and surface dyslexia was no longer trusted, because the patients that provided the data were later seen as “unpure” cases with other neurological co-morbidities, which eventually influences the results Marshall and Newcombe (1973) presented. Further patients with neurological disorders and lesions appeared, that showed different patterns of deficits, obviating the original distinction of the lexical and non-lexical module. However, as Shallice (1988) wrote, pure cases can only be identified with theories of the reading process that decide a priori, whether a case satisfies the criteria for being pure. So a priori assumptions of the existence of two routes/modules guides, which cases are seen as relevant, and which ones as non-relevant for testing the hypothesis that these two routes/modules exist. By adopting different assumptions, a pure case may always be re-described as impure (i.e., and exception case for which dual-route theory effectively does not apply in the sense that its predictions cannot be meaningfully tested with this case). Reviewing the progress of research on

neurological impairments and dyslexia with regards to dual-route theory for the following two decades, Van Orden et al. (2001) summarize that none of the failures to find consistent evidence for the two modules resulted in an effective falsification of dual-route theory, because its proponents argued that the evidence so far was flawed, consisting of impure cases. At the same time, ad-hoc auxiliary hypotheses (i.e., exceptions) were formulated as explanations for how the impure cases, even though not providing evidence for the two modules, would be entirely consistent with the theory, greatly complexifying the explanation of which factors and components are relevant for reading.

Here, we are back at the problem of interaction effects: Statistical interactions between component factors can thus not only occur within a single study, but also across studies. Nothing else is the continued appearance of new findings described by Van Orden et al. (2001) that did not fit a simple, two-factor model that originally could have been the dual-route theory of reading. Accordingly, for each interaction effect, an extra component or exception has to be formulated as an explanation for the results, which is added to the overall theoretical explanation of how reading works.

The question is now, where and when to stop? Of course, at some point in the future, the investigation of new factors/patients might lead to convergence on a finite set of findings that explain reading entirely. But this might just as well not be the case. Either way, the path along which those findings were generated resulted in an immensely complex set of statements that would need to be taken into account in order to formulate predictions, if a given case of reading or reading disorder is going to be investigated. The question of whether this can be seen as scientific progress has a strong practical component, namely whether the sets of statements is small to be testable. And whether the ad-hoc explanations generated are really scientifically productive, or merely constitute an ever-growing set of protective statements, what aim at saving the core hypothesis, even in the face of overwhelming evidence against it (Lakatos 1970)?

What this example should illustrate, is that progress and non-convergence (i.e., continued failure to observe consistent findings) lie very close to each other in the CDD paradigm. The problem is to tell the two apart: The observation of interaction effects themselves can be evidence for either development, and in the end, it becomes a practical or intuitive question for individual researchers again to judge, on which trajectory their research is heading. Hence, the observation of continued interaction effects by itself is not strongly conclusive. However, from the data analysis side based on the general linear model, there are further requirements as to what it takes to validly and reliably extract component effects, which is the assumption of independent and identical distribution of sources (iid-assumption) that compose the collected data from which component effects should be extracted.

4 1/f noise

General-linear-model type statistics assume independence of the collected data points, on which the statistics are then applied to draw conclusions. If data points are not independent, then model assumptions are violated, which can lead to faulty results. Violation of independence can be particularly well observed in timeseries data. Timeseries simply means, that multiple data points of the same observable are recorded from the same participant over time, so that a temporal sequence of the data points exists. However, the sequence of data collection should not influence the results, because otherwise the findings are not generalizable—they depend on how long or at what point in time or in which order of conditions the observable was measured. Of course, this can itself be made the object of investigation, but then we need to have the next time-invariance assumption, that the effect over time is always the same.

If we find an effect of time, and but the effect is constant (for example, such as when a value measured at time t is always twice the magnitude of a value measured at $t - 1$), then we can simply remove this effect from the data, and afterwards resume usual statistical treatment.

However, some data sets have been shown to exhibit correlation over time that are called “1/f noise”. The term 1/f noise derives from the spectral property of such time series, namely that when the timeseries is displayed as a power–frequency spectrum, its power drops off toward the higher frequencies as an inverse of the frequency f . If a time series contains 1/f noise, this means that data points are correlated over the whole observation period, so that every data point is connected to every other. The first article to report 1/f noise in time series of human behavior was published by Gilden and colleagues (Gilden et al. 1995), showing that consecutive intervals of time estimation (i.e., participants tapping their finger to a time-interval, such as 1 s) were correlated throughout the whole range of data points. This was an unexpected finding. Up to that point, psychologists assumed that the different trials in a task are independent—or near-independent. What this finding showed was that behavior was dynamically organized, so that a simple tap of a finger was not just a local event that unfolded on only within a limited time-scale of a few hundred milliseconds, but that other time-scales co-determined this behavior: The event happened in the context of behavior that occurred seconds or minutes ago.

Clearly, such data points were not independent. Moreover, this observation seemed to violate Simon’s (1973) proposal of near-decomposability of time-scales, where behavior and cognition of an organism is confined to a specific (or limited range of) time-scale(s), and can thus be separately analyzed from other events on other time-scales. Making the link from interactions of cognitive components (or at least experimental manipulations) in studies of human behavior to interactions of cognition and behavior across many time-scales, Van Orden et al. (2003) first proposed that the observation of 1/f noise in human behavior was clear evidence that the architecture of the cognitive system is not one of a finite set of independent components that sum together to produce human cognition. Instead of the

observation of continued interaction effects, which by itself did not yield solid evidence for a failure of a component-dominant explanation of cognition, observations of $1/f$ noise seemed to provide a definite test.

Moreover, they pointed to an alternative architecture of the cognitive system: Interaction-dominant dynamics. $1/f$ noise had been previously observed in physical systems in critical states during phase-transitions and in multi-component systems (Jensen 1998). Per Bak, a physicist working on the dynamics of such multi-component systems (such as piles of sand and rice) synthesized these two observations to propose that organisms work on the principles of self-organized criticality (Bak 1996): They are systems made up of many components (such as the human body), whose components are in permanent state of criticality. This means, that there is no fixed *internal* organization of the components (such as mental faculties encapsulated in the head and described by a psychological construct) that determines the behavior of the organism, but rather the behavior of each component is highly connected to the behavior of the other components, and a change in one of the components can quickly propagate through the whole organism to change the behavior of the other components. They are not independent of each other, but they interact.

This propagation of behavior of individual components can quickly lead to new organizations of the components, so that the many-component system can quickly and flexibly assume new states in order to adapt its behavior to a task or environment. Hence, the interactions between components “dominate” the behavior of the system, and no set of individual components does. Finally, physical experiments showed that $1/f$ noise was a hallmark of such interaction-dominant systems, its behavior characteristically exhibiting such fluctuations.

This provided a new concept of the cognitive architecture of an organism, one where the behavior was not dictated by a few mental components with specific ability (i.e., to perceive, compute, or control something), but where on a lower level relatively unspecific components interact to give rise to structured behavior that—measured in experiments with statistics that aimed at extracting independent components from such behavior—could be described as the result of a few dominant components controlling that behavior. However, the underlying architecture was potentially much less stable than a conception of hard-assembled set of components suggested.

From the view-point of interaction-dominant dynamics, changes in task contexts/experimental manipulations did not just change the behavior of one of the existing components, but potentially lead to a holistic reorganization of the cognitive-component system to adapt to the demands of the tasks. This, of course, would produce statistical interaction effects with a much higher probability than expected from a hard-wired system and might suggest the presence of a new component that was previously not observed from the view of CDD.

However, this kind of evidence—and hence the following conclusions—have been contested as well: Wagenmakers et al. (2004, 2005) noted, that other solutions to the inference problem to the observation of $1/f$ noise have been proposed as well. Here, we need to take again the perspective of an empirical scientist, who has two competing alternatives of how to interpret observations. Observing continued

interaction effects can be interpreted as a failure of research studies to yield consistent and stable results, or as a form of progress, where more and more is being learned about the phenomenon of interested.

The observation of $1/f$ noise in its relation to the former would suggest that the human mind is indeed a system made up of multiple, rather unspecific components, that, when put into a particular environment with a particular task, soft-assembles to a fitting component architecture. But this component architecture is emergent with regard to environment and task. Hence, observed “components” of this architecture are not basic and stably properties of the mind, and interaction effects between tasks do not actually mean that new additional information towards a more *complete* picture of the mind has been uncovered, but merely mean that the component architecture of the mind has adaptively changed from one set of task circumstances to another.

The question then is, how good evidence is $1/f$ noise for interaction-dominant dynamics? As Wagenmakers et al. (2004, 2005) noted, $1/f$ noise—also termed long memory—has also been observed in other disciplines, among them econometrics. Here, however, it has not spurred a debate about the basic architecture of economies, but rather has been treated as a statistical problem to deal with: Granger and Joyeux (1980) explained that $1/f$ noise (a.k.a. long memory) is a kind of nonstationarity that needs to be dealt with before the data—asset prices—that exhibit such fluctuations can be further analyzed statistically. Hence, Granger and Joyeux (1980) invented a method to remove such autocorrelations in time-series, called the back-shift operator, basically a filtering procedure.

As Kelty-Stephen and Wallot (2017) discussed, the implication of the availability of a method to statistically remove $1/f$ noise from an observable brought with it the option to interpret it merely as kind of a statistical error term, and not as unique evidence for interaction-dominant dynamics as a principally different way of how the human mind works. The evidence was now ambiguous. However, the application of the back-shift operator was a clever bit of modeling ingenuity that inadvertently opened up a theoretical quandary for interpreting long-memory in terms of a component-dominant view of the mind, and created direct problems for modeling such data with regard to the statistical degrees-of-freedom available for constructing models of such data (Kelty-Stephen and Wallot 2017).

Nevertheless, research on $1/f$ noise in behavioral measurements led to other observations that yielded more solid evidence in favor of an interaction-dominant interpretation, and this had to do with the time-dependence of $1/f$ noise patterns.

5 Multiplicative Interactions Across Scales

In 2010, Ihlen and Vereijken published a seminal paper entitled “Interaction-dominant dynamics in human cognition: beyond $1/f^\alpha$ fluctuation”, which was essentially a re-analysis of many of the data sets that had been used to fight over the presence, correct quantification, and interpretation of $1/f$ noise in human behavior. The essence of the paper was the observation, that most data sets that had been used in this argument—on both sides—exhibited what is called multi-fractal structure.

Fractal structure or more correctly, mono-fractal structure simply constitutes the observation that there is a power-law scaling relations between two quantities in a time-series (or spatial layout). Less simple, it also constitutes the assertion that this power-law structure is the result of a recursive process across multiple spatio-temporal scales. In any case, it has been used mostly used as another synonym for $1/f$ noise or long-memory, where quantities in the observed data exhibit one stable scaling relation.

Multi-fractal, as the name implies, means that multiple scaling relationships in a single data set are observed, and in the case of Ihlen and Vereijken's (2010) study, this meant multiple such relationships over time within the same timeseries of behavioral data set. This posed a problem for the backshift-operator solution proposed by Granger and Joyeux (1980), as this solution did not permit the filtering of a time-dependent $1/f$ pattern. This means, again from the perspective of an empirical researcher, weighting the evidence for a particular interpretation of the underlying architecture of the cognitive systems, this pattern constituted more unique evidence for interaction dominant dynamics. While a component-dominant treatment and interpretation of multifractal fluctuations is not available to date (see Kelty-Stephen and Wallot 2017), there are models that suggest something about the origin of such patterns in an observable, and they fit very well to an interaction-dominant interpretation.

A model that generates artificial multi-fractal data is the multifractal cascade process (Kelty-Stephen et al. 2013). In this process a multi-fractal timeseries is generated by a seed-value that is multiplied by sequence of values situated on different "scales", until the scale of the actual to-be-generated series is arrived at. This implies two important things: First, that a single value in the time-series does not (solely) stand for a particular instance that of something else that happened at this point in time, but rather is (co-)determined by events on higher time-scales (i.e., has common multipliers with neighboring events). Second, that the different scales that constitute the data are not independent (i.e., separate) of each other, but interact. Removing one of the scales (e.g., setting all multipliers of the scale to zero or eliminating them otherwise) will not yield results in the generation of a series.

This stands in contrast to a component-dominant view, which necessitates separable scales, as described by Simon (1973), and rather aligns with an interaction-dominant interpretation, which emphasizes that observed timeseries data of human behavior are an outcome of multiple interacting components across scales (i.e., potentially down from the time-scale of cellular processes to-and-above the time scales of actual task within which human activity is measured and observed).

6 Interaction-Dominant Dynamics

Component-dominant dynamics suggest a causal architecture, where the mind is made up of a set of different independent components/modules that possess certain "intelligences" (Turvey and Carello 1981), such as perceiving, classifying, remembering, retrieving etc. Experiments and the use of linear modelling, which allows to extract and describe independent sources of variation in observables, allow

to investigate such a set of components. At the end, one would ideally possess a list of components (and their connections), that represents a full theory of the mind. However, as we have discussed above, the requirements are that the different components work independently of each other, or more precisely, that they can be independently manipulated in experimental studies, and identified as independent sources of variation in relevant observables (Donders 1869; Sternberg 1969).

A prerequisite for this is, that these components work on a confined time scale, and that their activity on this time-scale is independent (“near-decomposable”—Simon 1973) of other time-scales. Otherwise, an assignment of variation in an observable to one source becomes difficult, and the distinction between different intra-organismic, as well as intra- and extra-organismic contributions becomes difficult, because they are now potentially conflated in joint time-scales of activities.

In contrast, interaction-dominant dynamics suggest a causal architecture of the mind where the components are potentially extremely numerous, but definitely not independent or characterized by possessing certain distinct intelligence. Instead, the rather unspecific components are strongly interactive, and self-organize to functionally adapt to a particular task and environment. However, this implies that the role the components play in a particular situation—such as an experimental task—is not defined by the intrinsic structure of these components (i.e., encapsulated mental faculties), but are (co-)defined by the task and environment, as well as the activity of the other components. Hence, all the time-scale that are involved in a task (i.e., from the beginning to the end of an experimental observation and potentially beyond, insofar as prior learning effects of the acquisition of particular skills plays a role for task performance) co-determine what is measured on the faster time-scales sampled within this task. The time-scales are *interdependent*. Hence, the observation of particular components via variations of the additive-factors-logic (Sternberg 1969) and linear modelling do not tap into the task- and time-invariant building blocks of the mind, but rather describe an emergent structure that can change over tasks, task-environments, and with time. Accordingly, these components are not stable, which can lead to a proliferation of interaction effects, for example, which in turn are the statistical symptoms of problems replicability and generalization of results.

To summarize: Originally, the interaction-dominant interpretation was evidenced by the observation of $1/f$ noise in behavioral (and neuro-physiological) data. An analogy was drawn to simple physical systems (Bak 1996; Jensen 1998), which were constituted by many interacting components, and produced emergent behavior that could be interpreted as the activity of a limited set of components—given certain constraints on larger time-scales. The observation of multi-fractal fluctuations expanded this picture. On the one hand, it lent more solid evidence to the interpretation of human behavior to be interaction-dominant compared to the mere observation of $1/f$ noise. On the other hand, it also suggested in more detail how the assumption of separability of different time-scales is violated, and transitions between contexts can lead to qualitatively different dynamics (i.e., interaction effects).

Viewed from the multi-fractal cascade model, we see that the generation of a multi-fractal timeseries necessitates the dependence of contributions from different

time-scales in a way that no single time-scale can be removed. Moreover, the cascading structure illustrates simultaneously that potentially all prior events could have an influence on a current performance, for example during a particular experimental task. At the same time, scale-breaks—such as when switching from one task to the next—can be reflected in changes in the dynamics of behavior by changes in the cascading structure at that specific scale. This is also conceptually an important step, because the classical $1/f$ case usually assumed a totally homogenous system (such as piles of sand or rice), which the human organism clearly is not. On the other hand, observations of changes in $1/f$ noise were hard to explain in this context, because a particular scaling relation was assumed to be an “ideal” marker of interaction-dominant dynamics, which should be equally present under all conditions (e.g., Goldberger et al. 2002). The last word in this debate is surely not yet said. However, given that there is tangible evidence for an interaction-dominant architecture of the mind, what are its consequences for scientific practice?

7 Consequences of Interaction-Dominant Dynamics for Generalization and Replicability

The practical side of “doing science”, given an interaction-dominant architecture of the mind—or, to be more precise, given a statistical structure of recorded data that does not conform to the requirements of CDD, but rather IDD—has so far not received much thought in the current debates. Obviously, one first steps has been to use new statistical procedures to investigate the presence of mono- and multi-fractal structure in behavioral measurements—in addition to, or at expense of the standard statistical analysis. The result has clearly been that such structure is quite pervasive, and by now, many studies have shown experimental manipulations of mono- and multi-fractal structure (e.g., Anastas et al. 2011; He 2011; Holden et al. 2011; Kerster et al. 2016; Kuznetsov and Wallot 2011; Stephen et al. 2012; Wallot et al. 2015a, b, c). However, it is not entirely clear how to proceed with such results. Some studies have tried to interpret fractal structure in terms of a more concrete property of the mind, somewhat like a kind of component (e.g., Correll 2008; O'Brien et al. 2014; Wijnants et al. 2012), or in terms of mind-environment relationship, somewhat like a more specific mechanism (e.g., Kloos and Van Orden 2010; Van Orden et al. 2011). Obviously, such efforts are not wholly consistent with the radical version of interaction-dominant dynamics that such measures are thought to index. Moreover, it seems unlikely that all differences in human behavior and thinking can be sensibly reduced to a single dimension of complexity, such a fractal structure. Using such measures, however, is surely a necessary starting point.

Another problem is the experimental task as such: Especially experiments that contrast distinct groups of treatments with a between-participant design do not allow to trace effects of time or influences of task-changes that could lead to different interpretations of similar tasks (see the problem described in “Interaction effects and the general linear model”). The first step would be to use repeated measures designs more frequently, and to either use them in a way that one can observe the unfolding performance to investigate when and where a re-organization of the

cognitive systems takes place, and what this means for putative cognitive components that participate in such tasks (e.g., Stephen et al. 2009)—or to continuously manipulate independent variables in an experiment, in order to observe whether this continuous manipulation actually leads to continuous effects, of qualitative changes in task performance, that are indicative of a change of the component-architecture (Kelso 1995).

Here, we also want point out that the observation of multifractality does by no means hinge on the use of repeated-measures designs where each participant performs all of the conditions. The initial observations of multifractality came from individual records of between-participants designs (Ihlen and Vereijken 2010), and to our knowledge, all studies that have investigated multifractality in behavior and (neuro)physiological data that were based on between-participant designs have found multifractal scaling (e.g., Booth et al. 2016; Carver and Kelty-Stephen 2017; Dixon and Kelty-Stephen 2012; Eddy and Kelty-Stephen 2015; Kelty-Stephen and Dixon 2014; O'Brien et al. 2014; O'Brien and Wallot 2016; Teng et al. 2016; Wallot et al. 2014; Wijnants et al. 2012). Also, collecting very short timeseries or between-participants data does not eliminate the problem of multifractality. Only because a timeseries is too small to reliably quantify multifractality, or because multifractality in one condition cannot be directly related to multifractality in a preceding condition with a participant does not mean that the underlying problem (instability and re-organization of the cognitive architecture) is gone. It is merely means that potential origins of nonstationarity on the level of observation or latent variables cannot be traced anymore.

However, not all research questions can be addressed in within-participant designs (e.g., research where learning effects during first exposure are important, applied research—for example on psychotherapy), not all manipulations can be operationalized in terms of continuous incremental changes (e.g., research employing complex action-sequences, research contrasting qualitatively different situations), and not all studies can be conducted in a way to collect a sufficient number of data points to conduct fractal analysis (e.g., questionnaire research, observational research, research employing invasive measurement procedures...). For some of these cases, effectively relying on the CDD logic (at least its measurement and analysis part) might be inevitable. For other cases, some workarounds might be to run different experimental conditions back-to-back, and continue measuring across the period where one task end and the new task (or experimental condition begins). The transition between two experimental conditions can be similarly informative as to whether these two conditions are comparable, or lead to a re-organization of the cognitive system (Wallot 2014).

The logic behind this suggestion is, that testing for IDD-characteristics can be informative with regard to the boundary conditions for which a CDD-description might be stable. Say we want to compare two experimental conditions, A and B, and want to do so in a classical CDD-style of analysis and explanation, then testing for IDD-characteristics of the collected data in A and B can yield evidence whether the comparison of those two conditions is permissible—form a pure statistical standpoint. In a most ideal case, the data collected in A and the data collected in B both show no signs of complexity characteristics (i.e., the data is iid-gaussian

noise, and no significant amount of (multi-)fractal structure is observed), and there are no nonlinear-transitions in the performance when changing from task A to B (i.e., no signs of re-organization of the cognitive systems as a function of change in task context). Then, a standard linear comparison of the two conditions seems warranted, we might have good reason to assume that the same boundary conditions hold and a CDD-solution might be stable. In a less-ideal case, we would not find evidence for observe multifractal structure and no evidence for phase-transitions between A and B, but the same magnitude of monofractal structure. Here, a CDD-solution might also be stable, even though the error terms of the statistical model will be higher than what would be expected from iid-gaussian noise, depending on the treatment of the data. However, evidence of multifractality within a condition, or evidence for phase-transitions between conditions suggests that stability assumptions are not met, and boundary conditions for a CDD-solution might not hold.

For situations where timeseries measurement is principally possible, but not at the order of several hundred data points, other complexity properties than fractal scaling can be investigated, which are indicative of IDD and suitable for shorter timeseries (e.g., Schiepek and Strunk 2010; Webber and Zbilut 1994), but whose relationship to fractal scaling has yet to be clarified.

However, a major problem that is not tackled by any of these approaches is the time-scale and task limitations of experimental setups in general. Experimental observations are always limited in terms of the time-scale that they are performed over with regard to the actual phenomenon of interest that they should bring under control or investigation. The possibility of substantial time-dependent changes of the cognitive architecture that interaction-dominant dynamics afford, however, implies that experiments are thus fundamentally limited in terms of the question of generalization. This problem is exacerbated by the fact that in an interaction-dominant causation there are potentially interactions between task and the way the cognitive systems is organized within the given task—and that interaction effect could be time-dependent as well.

This leads us to the principle problem of generalizations based on experimental observations. Admittedly this is a proposal, but it seems that the current practice in psychology and related sciences is to assume by default that findings in one experiment—or a series of experiments—will generalize, and that the surprising case is a finding that they do not. If the cognitive system is stable in the component sense assumed by Donders (1986) and Sternberg (1969), this seems justified. Cognitive components are independent of other components and extra-mental or extra-organismic influences, and once they have been successfully identified through proper experimentation, their presence should be assumed as given and stable. If, however, such components are not fundamental building blocks of the cognitive system, but emerge on the bases of interaction-dominant dynamics, and their emerging is contingent on other pre-conditions on other (time-)scales rooted in the past and/or the task environment, a new task environment might change the cognitive architecture and thus undermine predictions transferred from one task to another based on the assumed generalization of findings on the level of cognitive components.

In the absence of a relatively specific theory that predicts across tasks which findings will generalize or not—that is, in the absence of knowing the boundary conditions of observed “basic” effects, generalization cannot be assumed. Rather, the current implicit logic needs to be turned on its head, where findings in one experiment cannot be assumed to generalize, and investigation of putatively the same phenomenon in a different context first needs to establish that this new (experimental) context is similar in a way that generalization is possible to begin with. However, work on the boundary conditions of effects or principle cognitive components is rare to absent in current research. This can only be improved by carefully expanding experimental paradigms vertically, where a task in which certain effects have been found is only gradually changed at a time. Possibly, this also needs to be accompanied by horizontal expansions of research, where experimental effects need to be more systematically replicated in increasingly ecological settings, and observations from such settings need to be brought back into experimentation again. Otherwise, there might be a fundamental—not just gradual—disconnection between what is done in psychological laboratories and the real-world phenomena that are thought to be investigated in those laboratories (or, of course, already between different things that are done on the laboratory-level as well).

Finally, this kind of generalization problem due to interaction-dominant architecture also has implications for the problems currently discussed as the “replication crisis in psychology” (e.g., Maxwell et al. 2015). Replication of experimental findings also implies minimal requirements on the stability of the effects that ought to be replicated—and definitely requirements on the stability of the cognitive (component) architecture that is underlying these effects. If, however, the cognitive architecture is critically dependent on certain details of an experimental context—from rather “minor” aspects of the stimulus display (Van Orden et al. 2001, 2003) or larger cultural embedding of individual participants that are being tested (Henrich et al. 2010), as interaction-dominant dynamics suggest, then replication of an observation, just as generalization, cannot be taken for granted by default. While a more detail-oriented experimental task analysis and description might help some of the problems here, other factors are just much harder to equate and control—or to foresee in terms of their influence on performance.

Rather, one would need to understand what the architecture of the cognitive systems—and its activity—is dependent on under particular circumstances. But this needs to first acknowledge that such a relationship exists to begin with, and replace the assumption that cognitive components are independent, task-invariant, and basic building blocks of cognition.

To be sure, by pointing out the relation between replicability and an interaction-dominant architecture of the cognitive system, we do not mean to undermine the efforts that are currently being made to improve scientific practice. We hold it to be very likely that large parts of the problem of replication are due to sub-optimal scientific practice, such as “*p*-hacking” and publication bias towards “significant effects”, and by relatively “simple” statistical sampling problems, where an unlikely effect is being found by chance, then published, but effectively not re-observable (given the same statistical power). Nevertheless, we also want to point

out that overoptimistic assumptions about the stability of the cognitive architecture, such as implied by a component -dominant architecture, can contribute to this problem, and that there is evidence that these stability assumptions are not met. Related to generalization and replicability, this would also explain parts of the practice that Sullivan (2016a, b) observed behind the problem of construct stabilization: Researchers deem the cognitive systems and its architecture stable enough, so that widely different task and measurement situations are labeled as to investigate the same thing, and are actually also thought to tap into the same phenomenon, only because of the assumption of a component-dominant architecture of the cognitive system. Once a component has been discovered, so might go the logic, it is a basic part of the mind/brain and the stability of its contribution can be assumed under any variation of the actual task context. Hence, not greater care to align experimental task contexts is necessary. However, the observation of (multi-) fractal fluctuations in human behavior points to violations of the stability assumptions that are requirements for replicability and generalization in a CDD-oriented paradigm, and investigations of fractal properties provide a straightforward test of whether such stability assumptions are justified.

Acknowledgements Open access funding provided by Max Planck Society.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abraham, R. H. (1987). Dynamics and self-organization. In F. E. Yates (Ed.), *Self-organizing systems: The emergence of order* (pp. 599–613). New York: Plenum Press.
- Anastas, J. R., Stephen, D. G., & Dixon, J. A. (2011). The scaling behavior of hand motions reveals self-organization during an executive-function task. *Physica A*, *390*, 1539–1545.
- Bak, P. (1996). *How nature works: The science of self-organized criticality*. New York: Springer.
- Booth, C. R., Brown, H. L., Eason, E. G., Wallot, S., & Kelty-Stephen, D. G. (2016). Expectations on hierarchical scales of discourse: Multifractality predicts both short- and long-range effects of violating gender expectations in text reading. *Discourse Processes*. <https://doi.org/10.1080/0163853X.2016.1197811>.
- Brown, S. D., & Steyvers, M. (2009). Detecting and predicting change. *Cognitive Psychology*, *58*, 49–67.
- Carver, N. S., & Kelty-Stephen, D. G. (2017). Multifractality in individual honeybee behavior hints at colony-specific social cascades: Reanalysis of RFID data from five different colonies. *Physical Review E*, *95*, 022402.
- Correll, J. (2008). $1/f$ noise and effort on implicit measures of bias. *Journal of Personality and Social Psychology*, *94*, 48–59.
- Dixon, J. A., & Kelty-Stephen, D. G. (2012). Multi-scale interactions in *Dictyostelium discoideum* aggregation. *Physica A*, *391*, 6470–6483.
- Donders, F. C. (1869). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431.
- Eddy, C. L., & Kelty-Stephen, D. G. (2015). Nesting of focal within peripheral vision promotes interactions across nested time scales in head sway: Multifractal evidence from accelerometry during manual and walking-based Fitts tasks. *Ecological Psychology*, *27*, 43–67.

- Feyerabend, P. K. (2010/1975). *Against Method*. London, UK: Verso.
- Gilden, D. L., Thornton, T., & Mallon, M. W. (1995). 1/f noise in human cognition. *Science*, *267*, 1837–1839.
- Goldberger, A. L., Amaral, L. A., Hausdorff, J. M., Ivanov, P. C., Peng, C. K., & Stanley, H. E. (2002). Fractal dynamics in physiology: Alterations with disease and aging. *Proceedings of the National Academy of Sciences*, *99*, 2466–2472.
- Granger, C. W., & Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, *1*, 15–29.
- He, B. J. (2011). Scale-free properties of the functional magnetic resonance imaging signal during rest and task. *Journal of Neuroscience*, *31*, 13786–13795.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83.
- Hofstadter, D. R. (1981). Strange attractors: Mathematical patterns delicately poised between order and chaos. *Scientific American*, *245*, 22–43.
- Holden, J. G., Choi, I., Amazeen, P. G., & Van Orden, G. (2011). Fractal 1/f dynamics suggest entanglement of measurement and human performance. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 935–948.
- Ihlen, E. A., & Vereijken, B. (2010). Interaction-dominant dynamics in human cognition: Beyond 1/f² fluctuation. *Journal of Experimental Psychology: General*, *139*, 436–463.
- Jensen, H. J. (1998). *Self-organized criticality*. Cambridge: Cambridge University Press.
- Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge: MIT Press.
- Kelty-Stephen, D. G., & Dixon, J. A. (2014). Interwoven fluctuations in intermodal perception: Fractality in head-sway supports the use of visual feedback in haptic perceptual judgments by manual wielding. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 2289–2309.
- Kelty-Stephen, D. G., Palatinus, K., Saltzman, E., & Dixon, J. A. (2013). A tutorial on multifractality, cascades, and interactivity for empirical time series in ecological science. *Ecological Psychology*, *25*, 1–62.
- Kelty-Stephen, D. G., & Wallot, S. (2017). Multifractality versus (mono) fractality in the narrative about nonlinear interactions across time scales: Disentangling the belief in nonlinearity from the diagnosis of nonlinearity. *Ecological Psychology*, *29*, 259–299.
- Kerster, B. E., Rhodes, T., & Kello, C. T. (2016). Spatial memory in foraging games. *Cognition*, *148*, 85–96.
- Kloos, H., & Van Orden, G. (2010). Voluntary behavior in cognitive and motor tasks. *Mind and Matter*, *8*, 19–43.
- Kuznetsov, N., & Wallot, S. (2011). Effects of accuracy feedback on fractal characteristics of time estimation. *Frontiers in Neuroscience*, *5*, 62.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–195). London: Cambridge University Press.
- Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research*, *2*, 175–199.
- Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist*, *70*, 487–498.
- O’Brien, B. A., Wallot, S., Haussmann, A., & Kloos, H. (2014). Using complexity metrics to assess silent reading fluency: A cross-sectional study comparing oral and silent reading. *Scientific Studies of Reading*, *18*, 235–254.
- O’Brien, B. A., & Wallot, S. (2016). Silent reading fluency and comprehension in bilingual children. *Frontiers in Psychology*, *7*, 1265.
- Pinker, S. (1997). *How the mind works*. New York: Norton.
- Poldrack, R. (2010). Subtraction and beyond: The logic of experimental designs of neuroimaging. In S. Hanson & M. Buzzi (Eds.), *Foundational issues in human brain mapping* (pp. 147–159). Cambridge: MIT Press.
- Schiepek, G., & Strunk, G. (2010). The identification of critical fluctuations and phase transitions in short term and coarse-grained time series—a method for the real-time monitoring of human change processes. *Biological Cybernetics*, *102*, 197–207.
- Shallice, T. (1988). *From neuropsychology to mental structure*. New York: Cambridge University Press.

- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366.
- Simon, H. A. (1973). The organization of complex systems. In H. H. Pattee (Ed.), *Hierarchy theory: The challenge of complex systems* (pp. 1–27). New York: George Braziller.
- Stephen, D. G., Anastas, J. R., & Dixon, J. A. (2012). Scaling in executive control reflects multiplicative multifractal cascade dynamics. *Frontiers in Physiology*, *3*, 102.
- Stephen, D. G., Boncoddio, R. A., Magnuson, J. S., & Dixon, J. A. (2009). The dynamics of insight: Mathematical discovery as a phase transition. *Memory & Cognition*, *37*, 1132–1149.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276–315.
- Sullivan, J. (2016a). Construct stabilization and the unity of the mind-brain sciences. *Philosophy of Science*, *83*, 662–673.
- Sullivan, J. (2016b). Stabilizing constructs through collaboration across different research fields as a way to foster the integrative approach of the Research Domain Criteria (RDoC) Project. *Frontiers in Human Neuroscience*, *10*, 309.
- Teng, D. W., Eddy, C. L., & Kelty-Stephen, D. G. (2016). Non-visually-guided distance perception depends on matching torso fluctuations between training and test. *Attention, Perception, & Psychophysics*, *78*, 2320–2328.
- Turvey, M. T., & Carello, C. (1981). Cognition: The view from ecological realism. *Cognition*, *10*, 313–321.
- Van Orden, G., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, *132*, 331–350.
- Van Orden, G., & Kloos, H. (2003). The module mistake. *Cortex*, *39*, 164–166.
- Van Orden, G., Kloos, H., & Wallot, S. (2011). Living in the Pink: Intentionality, Wellbeing, and Complexity. In C. A. Hooker (Ed.), *Philosophy of Complex Systems. Handbook of the Philosophy of Science* (pp. 629–672). Amsterdam: Elsevier.
- Van Orden, G., & Paap, K. R. (1997). Functional neuroimages fail to discover pieces of mind in the parts of the brain. *Philosophy of Science*, *64*, 85–94.
- Van Orden, G., Pennington, B. F., & Stone, G. O. (2001). What do double dissociations prove? *Cognitive Science*, *25*, 111–172.
- Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation and interpretation of 1/f noise in human cognition. *Psychonomic Bulletin & Review*, *11*, 579–615.
- Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2005). Human cognition and a pile of sand: A discussion on serial correlations and self-organized criticality. *Journal of Experimental Psychology: General*, *134*, 108–116.
- Wallot, S. (2014). From 'cracking the orthographic code' to 'playing with language': Toward a usage-based foundation of the reading process. *Frontiers in Psychology*, *5*, 891.
- Wallot, S., Coey, C. A., & Richardson, M. J. (2015a). Cue predictability changes scaling in eye-movement fluctuations. *Attention, Perception, & Psychophysics*, *77*, 2169–2180.
- Wallot, S., Coey, C. A., & Richardson, M. J. (2015b). Interdependence of saccadic and fixational fluctuations. *Ecological Psychology*, *27*, 68–86.
- Wallot, S., O'Brien, B., Coey, C. A., & Kelty-Stephen, D. (2015c). Power-law fluctuations in eye movements predict text comprehension during connected text reading. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 2583–2588). Austin, TX: Cognitive Science Society.
- Wallot, S., O'Brien, B. A., Haussmann, A., Kloos, A., & Lyby, M. S. (2014). The role of reading time complexity and reading speed in text comprehension. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *40*, 1745–1765.
- Webber, C. L., & Zbilut, J. P. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology*, *76*, 965–973.
- Wijnants, M. L., Hasselman, F., Cox, R. F. A., Bosman, A. M. T., & Van Orden, G. (2012). An interaction-dominant perspective on reading fluency and dyslexia. *Annals of dyslexia*, *62*, 100–119.