Peer Community Journal

Section: Psychology

Opinion / Perspective

Published
2026-08-14

Cite as
Wanja Wolff, Sebastian Gluth and Johannes Keyser (2026) *We don't care how much you sweat: An epistemic framework for behavioral and brain science laboratory infrastructure*, Peer Community Journal, 6: e77.

Correspondence
wanja.wolff@uni-hamburg.de

Peer-review
Peer reviewed and recommended by PCI Psychology,
https://doi.org/10.24072/pci.psych.100132

We don't care how much you sweat: An epistemic framework for behavioral and brain science laboratory infrastructure

Wanja Wolff [1], Sebastian Gluth [2], and Johannes Keyser [1,2]

Volume **6** (2026), article e77

https://doi.org/10.24072/pcjournal.765

Abstract

For decades, behavioral and brain research has advanced by isolating single variables or brain regions to study behavior and performance. However, it has become increasingly clear that reductionist methods struggle to capture the complex, dynamic, and context-dependent nature of human behavior. Developments in data analysis and artificial intelligence now enable unprecedented insights into complex datasets. Yet, while different strands of reform literature have advanced how we theorize, measure, and analyze, the infrastructural conditions under which data are generated, the laboratory, have received comparatively little conceptual attention. Here, devices are often siloed, proprietary, or limited to aggregated outputs, thereby constraining the questions that can be addressed. In this paper, we offer an epistemic framework for reasoning about laboratory infrastructure at the scale of the whole laboratory, understood not as a collection of individual instruments but as an integrated infrastructure in which multiple hardware systems co-exist, communicate, and jointly support the questions a research group can ask. Conceptually, we think of measurements in three epistemic layers: the surface layer (raw numeric outputs, e.g., from an electrodermal sensor), the proxy layer (physiological or behavioral subsystems, e.g., sweat gland activity), and the target layer (emergent phenomena or constructs such as working memory capacity, arousal or effort). These layers are epistemic in that they describe how meaning is inferred from signals and the type of losses and mismatches that can occur at or between them; e.g., when we study arousal or effort, we don't care about sweat gland activity per se, but rather as an imperfect intermediate proxy. We derive from these layers a practical rating scheme across seven infrastructure properties: signal digitization, signal fidelity, temporal alignment, real-time access, interoperability, transparency, and flexibility. Researchers can use these to evaluate and optimize their laboratory across modalities. Our framework complements existing modality-specific standards and community-developed integration tools by providing a shared decision logic at the scale of the lab as a whole.

[1]Dynamics of Human Performance Regulation Laboratory, Institute of Human Movement Science, University of Hamburg, Germany, [2]Cognitive Modelling & Decision Neuroscience Laboratory, Department of Psychology, University of Hamburg, Germany

Peer Community Journal is a member of the Centre Mersenne for Open Scientific Publishing
http://www.centre-mersenne.org/

e-ISSN 2804-3871

Web-published in collaboration with UGA Éditions

Introduction

Traditionally, the behavioral and brain sciences have advanced by employing deterministic, reductionist approaches (Krakauer et al., 2017; Naughton et al., 2024; Pessoa, 2022). Such approaches encompass efforts to localize region-to-function mapping in neuroscience, identification and optimization of single performance markers in exercise science, or, more broadly, to establish the impact that a single variable or manipulation has on an outcome (Pearl, 2009). While this has generated important insights and has vastly increased domain-specific knowledge, researchers have argued that such approaches are ill-equipped to account for the inherent complexity and dynamic nature of human behavior (McLean et al., 2025; Scharfen and Memmert, 2024; Schüler et al., 2025). It has been questioned if we can truly understand "interesting behaviors" (p.56) by studying them at the level of isolated systems (Pessoa, 2019), and researchers across disciplines have highlighted the limitations that are inherent to reductionist approaches (Krakauer et al., 2017).

To illustrate, it has long been a mainstay of neuroscientific enquiry to search for the behavioral and cognitive processes that specific brain regions are functionally relevant for. However, such attempts to modularize the brain have had limited success: regions often attributed with a single, canonical set of functions — such as the amygdala as the *emotion hub* — are also reliably involved in many other processes (Pessoa, 2008). Complicating matters further, neural activity observed during a task may reflect epiphenomenal or context-dependent processes rather than direct functional involvement of the studied brain region (Tremblay et al., 2023). Consequently, brain functions are now increasingly conceptualized as entangled, context-sensitive processes supported by distributed networks (Hunt and Hayden, 2017; Pessoa, 2022). A parallel development can be seen in sports and movement science. Here, complex outcomes, such as performance, have traditionally been decomposed into discrete components (e.g., speed or strength), often neglecting how these components collectively and dynamically modulate the outcome of interest (McLean et al., 2025). In contrast, contemporary accounts increasingly recognize performance as an emergent result of non-linear interactions across multiple physiological and psychological sub-systems, that unfold across different spatio-temporal scales (Balague et al., 2013). Public health research is a third example: interventions to promote physical activity behavior have often been informed by relatively static and linear theoretical frameworks, targeting specific factors that are deemed relevant to being more physically active (e.g., focusing on risk-perception to increase motivation to act, or teaching an efficient training method to get fit fast) (Noar and Zimmerman, 2005). Yet, despite substantial efforts, these interventions have had relatively little lasting effect, as global physical inactivity remains high (Hallal and Pratt, 2020; Strain et al., 2024). Echoing the other domains, Schüler et al. (2025) call for a complex systems perspective on physical activity, emphasizing the need to account for interacting physiological and psychological processes that unfold dynamically over time.

At their core, the same message emerges across different fields of research: to understand interesting behaviors, reductionist models do not suffice and should be complemented by approaches that embrace spatio-temporal complexity. This not only echoes calls to study behavior in the real world (Bishop, 2008; Shamay-Tsoory and Mendelsohn, 2019), but more fundamentally highlights the importance of making laboratory infrastructure more capable to study interesting behavior. At the core, this has implications for the data we collect and the ways we analyze them.

Data analysis outpaces data acquisition

In *The Transmitter*, Dyer and Richards argue that "to gain insight into complex neural data, we must move toward a data-driven regime, training large models on vast amounts of information" (Dyer and Richards, 2025). This sentiment reflects a broader trend: conceptual pushes toward embracing complexity have been matched by rapid progress in data analysis techniques. Advances in, as well as simplified access to, machine learning and artificial intelligence now promise unprecedented insights into the hidden structures of large, high-dimensional datasets, while community efforts have made large strides to standardize data organization and curation (e.g., Brain Imaging Data Structure, BIDS; Gorgolewski et al., 2016 or FAIR data; Wilkinson et al., 2016). Combined with open repositories to freely share and re-use data (e.g., OpenNeuro; Markiewicz et al., 2021), these infrastructures greatly facilitate data-driven exploits at scale. While still incomplete and not universally adopted, these developments have changed what can be achieved once data have been collected. Amidst these impressive advances in data analysis and the enthusiasm for large-scale, data-driven science, comparatively little attention has been paid to where the data themselves originate: the laboratory. Many of the tools used to collect behavioral, physiological, or neural data often remain siloed and constrained. For example, devices might hide raw signals behind proprietary software, only export aggregated summaries without providing realtime[1] access to the acquired data, or lack interfaces that allow synchronization and event marking across systems[2]. However, such features are integral for capturing data that satisfy conceptual and analytical calls for complexity. The result can be a structural bottleneck: researchers can be forced to ask narrower reductionist questions, because their laboratory infrastructure cannot accommodate more ambitious designs. That such bottlenecks are real, and that solutions to them are sought after, is evidenced by the emergence and popularity of community-developed tools that aim to address them. Examples of such tools are Lab Streaming Layer (Kothe et al., 2025) for cross-device synchronization or Lab Linking for cross-laboratory integration (Schultz et al., 2024); for a real-world example from our experience, see Box 1.

With this paper, our goal is to help unblock this potential bottleneck and to maximize what we can learn from the data we collect. To do so, we propose an epistemic framework that aids in how laboratory infrastructure can be thought of. This follows the rich tradition of constantly re-thinking research processes in the behavioral and brain sciences (Munafò et al., 2017). As various interlocking activities are crucial for progress in the field, reform literatures exist for different such activities. For example, measurements have been classically examined under the banner of construct validity (Cronbach and Meehl, 1955) and, more recently, through the lens of questionable measurement practices and precision (Flake and Fried, 2020; Nebe et al., 2023). Theory and explanation have seen renewed attention through theory construction methodology (Borsboom et al., 2021), productive accounts of explanation (van Dongen et al., 2025), and diagnosing the field's theory crisis (Eronen and Bringmann, 2021). The analysis activity has long been the focus of statistical-power and replication reform (Cohen, 1962; Piray, 2026; Vul et al., 2009). Specific measurement modalities have their own excellent community standards. For example, SPR guidelines for EEG/MEG (Keil et al., 2014), EEG-BIDS for data structuring and naming (Pernet et al., 2019), as well as tutorial literatures for event-related potentials (Luck, 2014) or for eye tracking (Nyström et al., 2025). What has received comparatively little systematic attention is a prior condition that all these activities share: the laboratory as a whole, which we define not just as a collection of individual instruments but as an integrated infrastructure in which multiple systems of hardware and software interoperate and jointly support the questions a research group can ask. This is where we position this paper in the broader literature: we offer an epistemic framework for thinking about that infrastructure.

To clarify our paper's aim, we are explicit about what is inside and what is outside our scope. Our focus is on the laboratory as a whole, and we discuss infrastructural considerations that enable a research group to generate the data that are presupposed in other parts of the research process

[1] To improve readability, we will use the term real-time to refer to measures that provide a seemingly continuous data stream in real-time. Technically, sampling rates and delays will make (all/most) measures near real-time at best.

[2] This fragmentation is not only technical but architectural: behavioral science laboratories often rely on specialized cubicles in which a modality is measured in isolation. This maximizes precision within a narrow domain by optimizing testing conditions (e.g., temperature, humidity) for the modality of interest. By trading-off the ability to capture complex interactions across systems against deterministic precision, such an approach can have the unwanted side effect of reinforcing reductionism. Recent developments towards more interactive facilities such as the Dyadic Interaction Platform may be pivotal to overcome such architectural limitations (Isbaner et al., 2025).

(e.g., at the level of theory, measurement, and analysis), and how today's infrastructure choices may constrain the type of research questions one can ask tomorrow. As phenomena (Bogen and Woodward, 1988; van Dongen et al., 2025) and constructs (Cronbach and Meehl, 1955) of interest are rarely studied with one method alone, but through the interplay of different sub-systems (Nebe et al., 2023), we focus on the properties that matter when multiple measurement systems need to work together. Conversely, our work is not a standards document for a specific modality. Excellent modality-specific guidelines exist (Keil et al., 2014; Nyström et al., 2025; Pernet et al., 2019), with each modality focusing on their specific idiosyncrasies. With this paper, our aim is to provide a broader canvas on which modality-specific contributions can be organized, compared and additional modality-specific guidelines can be built upon. Our goal is not to provide a tutorial on measurement theory, construct validity, statistical analysis, or considerations for experimental design. All of these are essential, all of these can draw upon a rich and evolving literature, and none of which can be substituted for by simply buying a good device. In the same vein, we have not written a buyer's guide and we do not recommend specific vendors, devices, numeric thresholds or processing pipelines. Such recommendations are highly dependent on modality-specific constraints and context that the respective expert communities are better positioned to propose and revise. Instead, we provide a decision logic that can help researchers optimize their laboratory as a whole. For this, we introduce the idea of three epistemic layers of measurement and a set of practical infrastructural considerations that allow consistent reasoning about infrastructural choices across modalities, and facilitate evaluation and optimization of one's infrastructure against a shared vocabulary.

An epistemic framework: or why we don't care how much you sweat

To fully acknowledge the importance of laboratory infrastructure, it helps to re-evaluate what a measurement represents in the behavioral and brain sciences. Even with an instrument that is considered the gold standard for assessing a certain construct or phenomenon, we rarely directly observe what we are truly interested in. Instead, we propose that a measurement can be decomposed into (at least) three epistemic layers (figure 1) and different types of losses can occur at each layer.

At the surface layer, we record the output of our instruments, typically as numbers (figure 1, top row). These range from numbers that represent conductance changes in skin electrodes, voltages from an electrocardiogram, pressure fluctuations on a strain gauge, pixel intensities in an MR image, to responses on a psychological questionnaire or response times in a behavioral experiment. These values are crisp numerical outputs that represent inputs to the sensor, but their apparent precision is naturally bounded by the properties of the instrument: Temporal and spatial resolution, as well as the granularity of digitization determine which numerical outputs can be obtained in principle. For example, if skin electrodes were only sampled once per second, one would observe stepwise jumps and drops instead of smooth changes in skin conductance. Likewise, if the numeric output values of the sensor progressed in coarse increments (e.g., 1 micro-siemens steps), this would obscure more nuanced changes in skin conductance. Crucially, these numbers are not themselves the physiological or behavioral subsystem of interest, nor the constructs that motivate research in the first place. Put bluntly, we are rarely interested in how much a participant's hand is sweating, but in what such changes signify, how they emerge, and how they might interact with other (sub-)systems (a point eloquently made for electrodermal measurement already six decades ago by Montagu and Coles, 1966). Thus, at the surface layer, a key question is how precisely our instruments can capture the spatio-temporal characteristics of the underlying proxy layer.

At the proxy layer, surface layer signals are interpreted as indicators of specific physiological or behavioral subsystems (figure 1, center row). For example, changes in skin conductance are interpreted as an index of sweat gland activity; voltage changes in an electrocardiogram are interpreted as an index of the heart's activity; pressure on a strain gauge is interpreted as muscular forces that are exerted on a device; and fluctuations in fMRI signals are interpreted as changes in local blood oxygenation as a result of energy-consuming neural activity. Importantly, the mapping of subsystem dynamics to the sensor output is rarely one-to-one but is often rather loosely linked and corrupted by deterministic and probabilistic intermediate steps (cf. left side of figure 1).

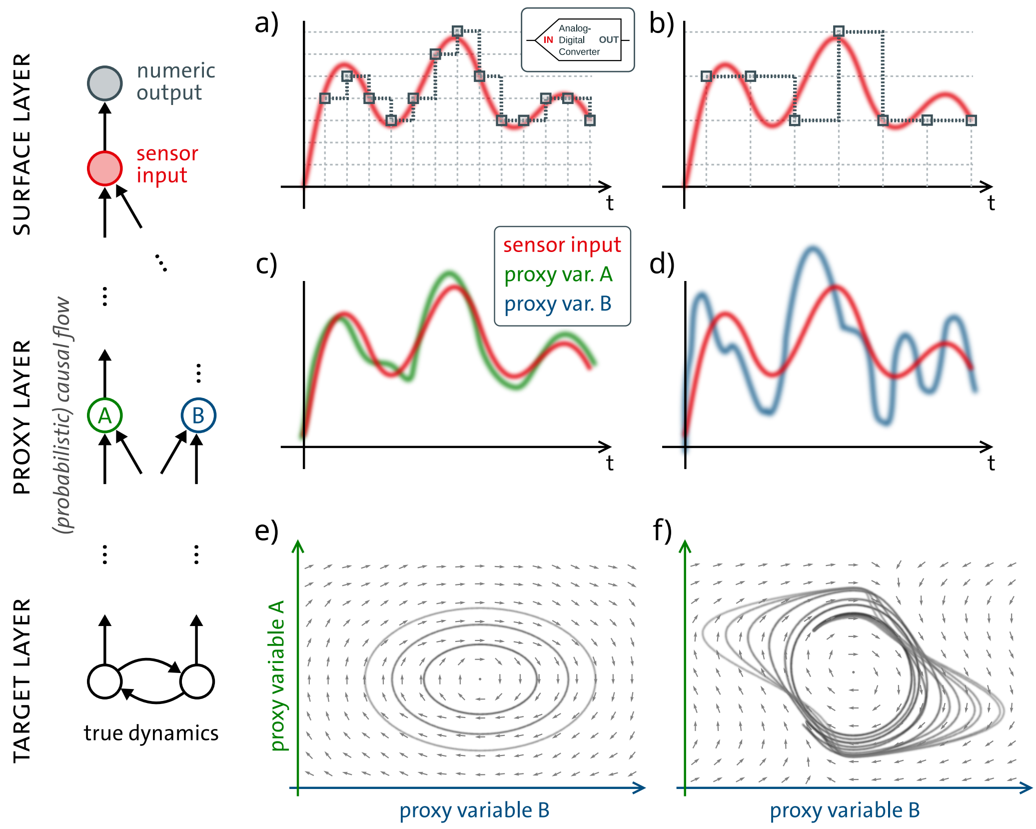

Figure 1 – Conceptual illustration of three epistemic layers of measurements. The layers are selections from the probabilistic causal flow starting at some (unknown) true dynamics of interest to a measured numerical outcome. On the surface layer, we obtain numeric recordings from a digital measurement device. Inside the device, the continuous sensor input is sampled into digital, numeric output. Different devices may sample the signal at higher (a) or lower (b) frequency and resolution (quantization precision). The intermediate proxy layer illustrates that the input into the recording device is generally just one choice among multiple possible physiological proxy subsystems, here labelled A and B. The true dynamics of the chosen proxy may be corrupted by other (noisy) causes before reaching the sensor input, thus reducing the fidelity between the actual recorded signal and the desired one, independent of recording device characteristics. To illustrate this, sensor input matches proxy A (c) much closer than proxy B (d). The target layer represents the goal of scientific inquiry that motivates the measurement. The unknown, true dynamics at this layer can only be observed indirectly through the other two layers. Generally, the target dynamics of interest are multi-dimensional, and thus allow the choice between different proxy dynamics that different types of devices may be able to measure. The chosen dynamics of interest range from simple (e) to complex (f) or even chaotic behavior, posing increasing demands on measurement and analysis techniques.
(Panel f shows the phase portrait of the Van der Pol oscillator, adapted from Wikimedia user Krishnavedala, licensed CC BY-SA 3.0).

Some fluctuations in the subsystem might not be measured by the sensor, for example, when a person inefficiently grips a handgrip dynamometer, much of the force the system generates does not reach the strain gauge and thus is lost. Conversely, nuisance factors that do not represent activity change in the subsystem of interest, such as temperature drifts, or motion artifacts, can cause a sensor to record a change that it should not. Mismatches between what the sensor records and the true subsystem dynamics can arise from a variety of other sources too, for example, from non-linear transduction (e.g., unequal sensitivity across contact points; Khalili et al., 2024), context (e.g., electrode placement, skin hydration; Goyal et al., 2022), or interference (e.g., skeletal muscle activity contaminating ECG signals; Wu et al., 2021). Crucially, these mismatches do not represent a lack of precision on the surface layer, because increased precision does not resolve them: increasing a dynamometer's sampling rate from 500 Hz to 1000 Hz does not make it a better measure of muscular force output, because the mismatch is not due to sampling rate but the imperfect mapping between strain gauge readings and the physiological subsystem of interest. At the proxy layer, the key question is therefore about fidelity: how closely does the sensor output reflect the true dynamics of the subsystem it is meant to capture?

At the target layer, we refer to the constructs and phenomena that motivate our research in the first place (figure 1, bottom row). This includes individual-level theoretical attributes such as working memory capacity, effort, arousal or trait anxiety, constructs in the sense of Cronbach and Meehl (1955). In addition, the target layer also reflects population-level regularities such as the Stroop effect or loss aversion, phenomena in the sense of van Dongen et al. (2025) or Bogen and Woodward (1988). At the level of theory and measurement, distinguishing between constructs and phenomena is integral. From a laboratory infrastructure perspective, however, the demands that target-layer entities place on how we record, synchronize, and integrate proxy signals are similar regardless of whether the ultimate interest is in a construct or a phenomenon. Most likely, constructs or phenomena are not fully captured through single subsystems, but are brought about by the dynamic interplay of different physiological and psychological processes that unfold over time in a context-specific fashion. For example, skin conductance is generally studied as a representation of sympathetic nervous system activity (Dawson et al., 2007), and a researcher might measure it to learn more about arousal (Lonsdorf et al., 2017). However, arousal itself is likely shaped by the dynamic interplay of neural, endocrine, cognitive, and situational factors. Similarly, muscular force might be studied as a representation of physical effort, but in its entirety, effort is far from being fully understood and a plethora of subsystems (and their associated measures) have been used to capture what effort truly is and what it does (Hack et al., 2025; Wolff et al., 2025). At this layer, a single sensor or proxy subsystem can rarely provide a complete answer. Thus, ideally, laboratory infrastructures allow researchers to flexibly and creatively combine various proxies and investigate their spatio-temporal interactions. Infrastructure that meets these requirements can facilitate experimental designs that go beyond mere observation towards interventions that can be applied and processed in real-time. For example, by manipulating one subsystem, such as providing biofeedback on heart rate, we can examine how perturbations cascade through other subsystems and modulate the dynamics of the construct of interest. In short, the target layer highlights why infrastructures that are interoperable, transparent, and flexible are essential: without them, our ability to study what we find truly interesting remains severely limited.

Naturally, the claim that measuring the things one finds interesting can be challenging is not new (Cronbach and Meehl, 1955; Nebe et al., 2023). Readers will likely recognize that our three layers bear resemblance to classical treatments of construct validity, which centers on the question if a test score can be interpreted as a measure of a construct of interest (Cronbach and Meehl, 1955). Likewise, recent work on theory-building differentiates verbal theories, formal models, statistical patterns, and phenomena (van Dongen et al., 2025). Our three layers are not a conceptual reorientation but a practical addition: where construct validity asks whether a test score justifies an interpretation, and productive explanation asks whether a theory can explain a phenomenon, our framework asks whether the laboratory can generate the data that either endeavor would require. This shifts the unit of analysis from the individual test or theory to the laboratory as a whole. To borrow Cronbach and Meehl's (1955) own example: Whereas they asked whether a measure of sweating validly indicates anxiety, we ask whether your laboratory infrastructure can capture the sweat response (or an alternative proxy) with sufficient precision, fidelity, and temporal alignment to answer their question.

Practical considerations for laboratory infrastructure: or why we really don't care how much you sweat

In the previous section, we detailed the challenges in measuring the things we are genuinely interested in. Surface level data, proxy dynamics, and emergent target constructs are connected through imperfect mappings, each layer imposing different challenges on researchers and manufacturers. Recognizing these epistemic layers clarifies why laboratory infrastructures matter, and why it can be beneficial to keep these mismatches in mind when setting up a lab in the behavioral and brain sciences. In the next section, we translate these epistemic layers into practical considerations for researchers seeking to future-proof their lab for the complex questions of tomorrow. We will not provide prescriptive rankings for specific devices, suppliers or measurement technologies. Instead, these considerations offer a conceptual vocabulary for evaluating and comparing the infrastructure choices that enable or constrain research on complex, dynamic behavior. As a practical starting point to guide decision-making, we also propose a ranking scheme (see table 1), similar to rankings in other domains, like tumor staging, the Nutri-Score for rating nutritional quality of food, or the various Tier lists in contemporary Internet culture that seem applicable to anything. Crucially, our drafted ranking scheme is based on the epistemic framework, flexible for adaptation by individual researchers and open to be shaped by the community as a whole.

Importantly, these considerations should not imply that each measure needs to excel in every dimension to satisfy one's research interests. In some cases, financial or practical constraints might offset the flexibility gains that are afforded by the most future-proof solution. Likewise, for some target constructs and research questions (e.g., reward-related brain activity in subcortical areas), even the best currently available technology does not fully satisfy any of the epistemic layers. Functional MRI, for instance, offers high spatial but poor temporal resolution, and although advances in acquisition speed and modeling can partially mitigate the sluggishness of the BOLD signal, it remains an inherent limitation of the method. Moreover, the numerical values obtained from an fMRI scan are no one-to-one representation of the physiological system being tracked (e.g., localized neuronal activity), nor of the research question one is trying to answer (e.g., how the brain processes rewards). Thus, epistemic transparency is key: being conscious of how one's laboratory infrastructure aligns with the type of questions one wants to study, and where mismatches (may) occur between the measured numerical output and the intended target constructs. The practical considerations outlined below can serve as guideposts for this decision process. Before detailing each category, we present a concrete example from our own experience to illustrate how these interact in a realistic research problem (Box 1).

Box 1: Bottlenecks due to siloed infrastructure can constrain the questions we can address. To illustrate how these considerations matter in practice, consider a study on the interplay between physical effort and attention. This question directly touches upon multiple infrastructure categories simultaneously. Our typical setup involves a stimulus computer for presenting stimuli and recording behavioral responses, a cycling ergometer for controlling and recording physical effort parameters, and a heart rate monitor. Already at the device-level, infrastructure choices define how this question can be studied. Cycling ergometers, for instance, differ in terms of precision and the information they provide (Bouillod et al., 2022). This constrains the effort dynamics that can be tracked at the surface layer. Heart rate measurements offer an illustrative case for constraints regarding signal fidelity and transparency. Many commercial heart rate monitors — whether wrist-worn devices based on photoplethysmography (Charlton et al., 2023) or chest straps based on electrocardiographic sensing (Machado et al., 2025) — transmit only aggregated beats-per-minute values to the ergometer. From this pre-processed summary, the underlying beat-to-beat dynamics cannot always be recovered, and the algorithm underlying the actual heart rate calculation may be opaque. For research questions that require heart rate variability or event-related cardiac responses, such a device would be insufficient at the proxy layer, regardless of how well the rest of the setup performs. A device that provides the raw electrical signal preserves the dynamics needed for more fine-grained analyses. These device-level constraints are compounded by how devices communicate. In many default configurations, the heart rate monitor sends its data to the ergometer that saves its data internally, resulting in two separate data files with no straightforward path to temporal alignment (figure 2, left). Even just enabling post-hoc alignment may require cumbersome workarounds to relay event triggers from the computer to the ergometer (figure 2, center). In our setup, for example, we rely on two additional signaling devices (one from a third-party vendor) and custom-built cabling to send simple event triggers from the computer to the ergometer. Although such a workaround can work very well, it may be fragile, time-costly, lab-specific, and difficult for others to replicate. An ideal configuration — in which the computer reads from and writes to all devices in real-time, enabling adaptive paradigms where stimulus presentation or ergometer resistance adjusts to the participant's current physiological state — requires interoperable software interfaces and middleware such as Lab Streaming Layer (figure 2, right). Notably, our illustrative design example here is a relatively simple one, where the need for cross-device interactions may already constrain which research questions are feasible. Naturally, this extends to designs that integrate additional devices, such as spirometry and ventilation data to capture physical effort dynamics at the proxy layer more completely, or an eye tracker to study how attention is deployed alongside behavioral responses. Each additional device multiplies the demands on temporal alignment, interoperability, and transparency.

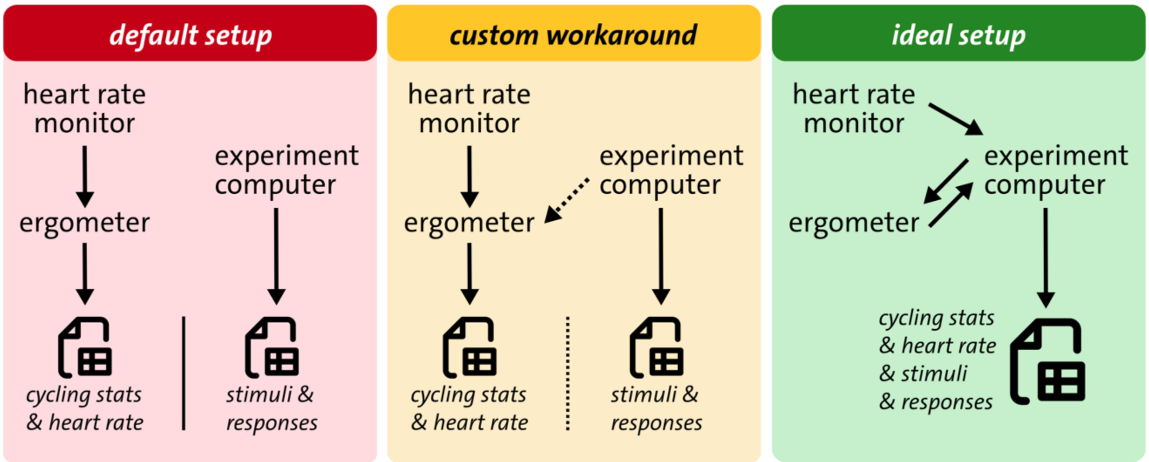

Figure 2 – Illustration of three setups involving a heart rate monitor to record heart rate, a cycling ergometer to evoke physical effort and to record cycling metrics, and an experiment computer to present stimuli and record responses. Left panel: In the default setup, the heart rate monitor sends its data to the ergometer, which saves them along with cycling metrics into an internal data file. The experiment computer saves data about stimuli and responses into a separate data file. Any fine-grained temporal alignment between the two files is impossible. Center panel: In a custom workaround, the computer sends event data to the ergometer, enabling post-hoc temporal alignment between both data files. Right panel: In an ideal setup, the computer directly receives the heart rate and cycling statistics in real-time, and is able to control the ergometer settings, allowing for adaptive paradigms and saving all temporally aligned data into a single file.

Signal digitization: Sufficient temporal and spatial resolution is required to accurately capture the fine-grained temporal dynamics that are inherent to many signals. Only adequate sampling rate and quantization resolution allows for analyzing dynamic interactions within and between subsystems (figure 1 a,b). For example, with handgrip dynamometers that track grip force at sampling rates of 100 Hz and higher, one might study intricate research questions on time-force relationships or the effects of subtle changes in the environment on force production. In contrast, low sampling rates or the mere provision of aggregated summary variables after a measure has been taken, constrain possible research questions already at the surface layer. To illustrate, with a handgrip dynamometer that only produces a force readout after it has been squeezed, one cannot analyze force production over time, nor how it is acutely affected by external factors. Importantly, higher resolution should not be equated with higher fidelity per se, as some to-be-measured subsystems might be inherently slow (thereby restricting research design flexibility on the proxy level and not the data acquisition level), and this cannot be overcome by ramping up the sampling rate. Nevertheless, sufficient resolution is necessary for the valid detection of patterns and transient phenomena that are characteristic of complex systems.[3]

Signal fidelity: Crucially, precision and fidelity are not the same. A device might record a signal with excellent numerical precision, but the numbers might reflect the underlying subsystem dynamics only loosely (figure 1 c,d). Here, the surface layer precision does not translate well to the proxy layer. Fidelity describes how well the recorded signal maps on the behavioral or physiological process of interest. For example, finger electrodes to measure activation changes in the autonomic nervous system can provide highly accurate read-outs of "how sweaty one's hands" are, but factors such as humidity can obscure the true dynamics (Bari et al., 2018). Similarly, some proxies require substantial pre-processing for a researcher to arrive at numerical outputs that ostensibly represent dynamics of the proxy of interest. For example, functional near-infrared spectroscopy (fNIRS) infers changes in oxy- and deoxyhemoglobin concentrations - which are interpreted as reflecting local hemodynamics - by tracking light absorption in the tissue of interest (e.g., brain, muscle). This process relies on a sequence of assumptions (e.g., regarding tissue composition), absorption laws (e.g., Beer-Lambert law), as well as a variety of data processing steps (e.g., filters, artifact correction algorithms, drift correction, regressing out contributions from external tissue) that transform the raw signal into a measure of oxygenation. While processing and reporting steps are continuously evaluated and optimized (Pinti et al., 2019; Yücel et al., 2021), they introduce potential sources of noise that can reduce the fidelity between how well the obtained signal represents the true physiological dynamics one is interested in (Tachtsidis and Scholkmann, 2016). Another challenge to signal fidelity is introduced by proxies (and the devices used to measure them) whose dynamics cannot straightforwardly be captured in real-time. For example, heart rate variability (HRV) is a popular measure of parasympathetic activity. It represents the variability between subsequent heart beats, mostly in milliseconds (Mosley and Laborde, 2024). While the true HRV is continuously changing, researchers need a series of heartbeats to reliably quantify HRV, thereby arriving at a more static numerical measure than the true dynamics that underlie the proxy one is interested in. As these examples show, even highly precise data acquisition does not necessarily lead to high signal fidelity as constraints of the device, or the proxy of interest can lead to mismatches. Understanding and documenting such mismatches is essential to ensure that conclusions drawn from a signal genuinely reflect the dynamics of the subsystem of interest. It is worthwhile to note that fidelity issues can arise from surface-level properties (e.g., non-linear transduction, temperature-drifts in the sensor), but they can also be caused by suboptimal mapping between proxy and target layers (e.g., too strict/loose signal processing choices, misguided interpretations like "a large GSR indicates that the person is lying"). The latter is best addressed through training (e.g., learning optimal processing) and our framework is primarily concerned with the former (e.g., buying the best-suited device to integrate into the broader infrastructure).

Temporal alignment & event marking: If one wants to study the dynamic interactions between multiple systems, temporal alignment of data streams is necessary. More specifically, complex experimental sequences (e.g., presentation of different stimuli that have to be classified according to different rules in a randomized within-subject design) or specific physiological or

[3]It is prudent to note that higher sampling rates come at the cost of heightened storage demands and may under certain conditions lower the signal-to-noise ratio of the acquired data. This can be a practical trade-off that researchers may need to weigh when making infrastructural choices.

behavioral events need to be automatically and precisely marked in each data stream, so researchers can pinpoint at what exact time a certain event had occurred and to study how this covaried with another measured subsystem. Devices that do not allow for automatic event marking substantially limit the complexity of experimental designs that can be realized. This is due to the high level of uncertainty that is introduced with manual (e.g., the experimenter presses a button or makes a note when a certain event has occurred) or post-hoc marking (e.g., by calculating at what time a certain event should have occurred in the experimental sequence and manually adding an event marker to the data file after its collection).

Real-time data access and interaction: If a recording device provides real-time access to the measured raw data stream, there is a qualitative jump in the type of feasible research questions, as it allows for tailored, adaptive interventions of the dynamics of interest. Real-time access to raw data allows researchers to visualize, process and provide feedback to participants during data acquisition. With such capabilities, one can, for example, study how participants integrate live feedback from one or more (bodily) systems during an experiment and how modulating the feedback signal (e.g., by providing false feedback in real-time) would alter how they feel and act. Such closed-loop capabilities expand the laboratory to a site where adaptive interventions can be studied. At first glance, these dynamic interactions are of particular interest in applied settings such as sports psychology and clinical psychology, in which real-time bio-feedback holds promise for improving performance and well-being (Fernández-Alvarez et al., 2022; Xiang et al., 2018). But they can also be highly desirable in more basic research domains, for instance, when carefully controlling a participant's access to visual information with an eye-tracker and a gaze-contingent stimulus presentation, so that peripheral viewing is disabled (e.g., Eum et al., 2023). In contrast, infrastructure that only allows restricted interactions or only provides post-hoc outcomes that prevent in-situ interactions restricts scientists to more static, coarse-grained research questions.

Interoperability & multi-modal integration: Interesting behavior is likely to emerge from the dynamic interplay of various subsystems. For example, a researcher might want to study if changes in visual attention are associated with event-related EEG potentials that signal belief updating processes in a decision-making task or vice versa. To study such questions, laboratory infrastructure needs to be set up in a way that allows for seamless communication between devices and modalities. This can be achieved if devices rely on (open) well-documented interfaces that allow fully customizable data streams whose outputs can interact with other concurrently running devices. Emerging solutions, such as Lab Streaming Layer (Kothe et al., 2025), hardware that specializes in synchronizing data streams across devices or simply devices that are shipped with a variety of physical in-/output ports can all facilitate interoperability, thereby paving the way for research questions that can tackle complex behaviors. In contrast, devices that prevent machine-to-machine interaction and that do not allow for real-time data access run the risk of producing epistemic silos that only provide answers to what one device can measure in isolation.

Transparency, metadata & documented processing: Transparency ensures traceability. Devices that provide access to raw data — or to processed data but with full documentation of the underlying processing pipelines — allow for reanalysis, replication, and reinterpretation as analytical tools evolve. Surprisingly, this seemingly straightforward desideratum is not often met as vendors (and probably customers too) prefer to provide data outputs that already represent the unit of measurement of the proxy the researcher is interested in. For example, an eye-tracking software might automatically compute specific events (saccades, fixations, blinks) from the raw x- and y-coordinates of the eye's position at every time point using an internal algorithm. Whereas most end-users would be content with this solution, it can prevent more fine-grained analysis such as saccade velocity or micro-saccade measures, and it can make integration with other tools such as EEG more difficult. When laboratories or vendors restrict access to raw data (or the underlying processing steps) or even restrict data export to proprietary data formats, they impose interpretive boundaries that cannot easily be revisited. Transparent infrastructures preserve the possibility of re-asking new questions of old data, which is a hallmark of sustainable science. Beyond raw data access, metadata should ideally conform to machine-readable community standards like BIDS and its modality-specific extensions (Gorgolewski et al., 2016; Pernet et al., 2019).

Table 1 – Rating scheme for laboratory infrastructure.

Category and Layer(s)	**A: Excellent**	**B: Medium**	**C: Limited**
Signal digitization (Surface)	Continuous recording at sampling rates and quantization appropriate to the fastest dynamics of interest.	Adequate for typical questions but limited room for faster or finer dynamics.	Sparse, summary-only, or fixed-rate output that cannot be reconfigured.
Signal fidelity (Proxy; with target-layer design implications)	Sensor output maps closely to the intended physiological or behavioral subsystem; known artifact sources are documented and can be mitigated.	Mapping is imperfect but well-characterized; major artifact sources can be modeled or corrected.	Mapping is unclear or loose; artifact sources are unquantified.
Temporal alignment & event marking (Surface-proxy bridge)	Automatic event markers with low timing uncertainty across all concurrent data streams.	Event marking with bounded latency; mostly accurate across devices.	Manual, post-hoc, or only approximate event logging.
Real-time data access & interaction (Proxy-target bridge)	Raw data stream is externally accessible with low latency, enabling closed-loop, adaptive designs.	Partial or delayed live access; supports simple monitoring but not adaptive paradigms.	Post-hoc access only; no support for interaction.
Interoperability & multi-modal integration (Target)	Open, well-documented interfaces; devices operate without mutual interference; straightforward combination with other modalities.	Integration is possible but requires custom or third-party engineering.	Closed, proprietary system; physical or technical incompatibilities constrain combination with other infrastructure.
Transparency, meta-data & documented processing (All layers)	Raw data accessible; all preprocessing steps documented or inspectable; metadata conform to community schemas (e.g., BIDS) where available.	Processed data accessible with partial documentation; metadata present but incomplete or non-standard.	Opaque processing; only derived outputs exposed; metadata missing or locked in proprietary formats.
Flexibility & re-use (Cross-cutting)	Modular hardware and software; components can be added, swapped, or repurposed.	Some reconfiguration possible but constrained by vendor ecosystem or fixed pipelines.	Single-purpose, closed design; difficult to adapt or extend.

For each category, it is indicated where in the surface-proxy-target chain the described category property primarily matters. Importantly, most categories likely have cross-layer consequences. The A/B/C ratings are intentionally qualitative and device-agnostic: calibrating them to concrete thresholds is the job of modality-specific communities. Please note, each category describes a property of the infrastructure, not of the analyses or the people operating it.

These inform other (and one's future self) about device specifications, sampling parameters, and processing provenance and increase chances that data remain interpretable and reusable beyond the laboratory where data have been collected. Of course, what constitutes "raw" data differs greatly across modalities and assessing true raw data may not always be feasible. To use an extreme example, image reconstruction from k-space data in fMRI requires highly specialized expertise that most cognitive neuroscientists do not possess. In such circumstances, optimal transparency would still require full documentation of all processing steps — as opposed to proprietary vendor-specific algorithms — so that their impact on the analyzed data can be evaluated and potentially revisited in the future.

Flexibility & re-use: Science moves forward, and we probably do not know which questions we want to answer in the future. Laboratory infrastructure is most valuable if it reflects this by remaining adaptable (technically, conceptually, and collaboratively). Naturally, this last desideratum is best met if the ones above are met as well. With flexibility, we refer to how well a setup can be reconfigured to address new research questions, whereas with compatibility for re-use, we refer to how easily it can integrate with other technologies or infrastructures (intra- and inter-lab collaboration; e.g., Schultz et al., 2024). The properties greatly affect how future-proof a lab is and how well it can evolve with scientific trends. This implies that laboratory infrastructure should ideally be modular in terms of the hardware and software it relies on, such as sensors and devices that do not "cancel each other out" (using x does not allow using y), open and flexible interfaces that allow for parallel acquisition of multiple signals with minimal cross-talk or interference, and modular acquisition systems that can scale in complexity as new components are added. Optimizing flexibility and compatibility maximizes the type of research questions one can sensibly tackle, extends the lifespan of the laboratory infrastructure, and makes interdisciplinary collaboration easier — within and across laboratories — as specialty knowledge and devices can be added and removed in a modular fashion. In contrast, infrastructure that is optimized towards one specific (type of) protocol or whose physical, electromagnetic or optical constraints make it incompatible with other tools limit the levels of analysis that can be realized. Examples of such constraints might be the movement restrictions that are inherent to fMRI, EEG or (non-OPM-)MEG, or the incompatibility between different technologies that rely on optical signals (e.g., combining certain types of optical motion capture technology with fNIRS).

Conclusion

In this paper, we argue that progress in the behavioral and brain sciences will increasingly depend on capturing complex dynamics across multiple time scales and levels of analysis. This challenge can hardly be met by focusing on a single measurement tool, no matter how precise and reliable this tool may be. Instead, a future-proof lab infrastructure will require the effective integration of multiple devices, which in turn requires considering their individual and collective capabilities to accurately measure the phenomena and constructs we are ultimately interested in. We propose that these considerations should move along three epistemic layers: whether a lab's devices provide sufficiently precise data points (surface layer) of sufficiently informative physiological or behavioral markers (proxy layer) to enable studying the complex interactions at the target layer. On the practical level, we derive seven infrastructure properties along which the capabilities of one's laboratory can be evaluated (table 1). Because it is the interplay between devices that ultimately determines what a lab can investigate, our unit of analysis is the lab as a whole, not any individual device. Excellent modality-specific standards and guidelines already exist and continue to evolve. Community-developed tools such as Lab Streaming Layer (Kothe et al., 2025) and Lab Linking (Schultz et al., 2024), as well as data-structuring and naming conventions like BIDS (Gorgolewski et al., 2016) and its modality-specific extensions (Pernet et al., 2019), address concrete facets of the integration problem we describe. These efforts and ours are sequential and complementary: community standards specify how data should be structured and documented once they exist. One function of our framework is to ask whether the laboratory infrastructure can produce data across devices that lend themselves to such structuring in the first place. Our contribution can therefore be seen as an epistemic canvas to connect these efforts: a shared vocabulary and decision logic that allows researchers to think about infrastructure choices consistently across modalities, and that helps identify where further community work is needed.

We recognize that optimal infrastructure will not automatically produce interesting research. Many challenges in the behavioral and brain sciences stem from suboptimal data acquisition, processing or faulty statistical analyses. These are issues that boil down to methodological expertise rather than hardware limitations (Flake and Fried, 2020; Munafò et al., 2017; Nebe et al., 2023). Rigorous training in how to work in the laboratory, how to design experiments, and how to handle data cannot be replaced by better infrastructure. Even the best infrastructure — according to our ranking scheme — is of limited use without the expertise to actually apply it. Our paper addresses the complementary problem: for well-trained researchers to maximize their potential and study the most interesting questions, material preconditions matter. Even if everyone in the laboratory is an expert methodologist, it will still be limited when devices cannot interoperate, or when raw data are inaccessible due to proprietary software. Realizing the potential of future-proof laboratories will therefore also require institutional investment in methodological training, dedicated laboratory staff, and evaluation criteria that reward the quality of research infrastructure alongside research output (Gärtner et al., 2025; Schönbrodt et al., 2025).

Importantly, we acknowledge and value the important contributions of laboratories that specialize in individual methods. Yet, it is our opinion that in order to fully understand "interesting behaviors", the canonical future lab will need to embrace integrating a variety of methods. We hope that adopting (and improving) our ranking scheme will help communication and decision-making for setting up the best possible lab infrastructure to study complex, dynamic behavior.

Acknowledgements

Preprint version 3 of this article has been peer-reviewed and recommended by Peer Community In Psychology (https://doi.org/10.24072/pci.psych.100132, Frischkorn, 2026).

Fundings

The authors declare that they have received no specific funding for the article.

Conflict of interest disclosure

The authors declare that they comply with the PCI rule of having no financial conflicts of interest in relation to the content of the article.

Data, script, code, and supplementary information availability

No data, scripts, code, or other supplementary information are associated with this article.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT and Claude in order to remove typos and improve readability of some sentences. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the article.

References

Balague N, Torrents C, Hristovski R, Davids K, Araùjo D (2013). Overview of Complex Systems in Sport. *Journal of Systems Science and Complexity* **26**, 4–13. https://doi.org/10.1007/s11424-013-2285-0.

Bari DS, Aldosky HYY, Tronstad C, Kalvøy H, Martinsen ØG (2018). Influence of Relative Humidity on Electrodermal Levels and Responses. *Skin Pharmacology and Physiology* **31**, 298–307. https://doi.org/10.1159/000492275.

Bishop D (2008). An Applied Research Model for the Sport Sciences. *Sports Medicine* **38**, 253–263. https://doi.org/10.2165/00007256-200838030-00005.

Bogen J, Woodward J (1988). Saving the Phenomena. *The Philosophical Review* **97**, 303–352. https://doi.org/10.2307/2185445. JSTOR: 2185445.

Borsboom D, van der Maas HLJ, Dalege J, Kievit RA, Haig BD (2021). Theory Construction Methodology: A Practical Framework for Building Theories in Psychology. *Perspectives on Psychological Science* **16**, 756–766. https://doi.org/10.1177/1745691620969647.

Bouillod A, Soto-Romero G, Grappe F, Bertucci W, Brunet E, Cassirame J (2022). Caveats and Recommendations to Assess the Validity and Reliability of Cycling Power Meters: A Systematic Scoping Review. *Sensors* **22**, 386. https://doi.org/10.3390/s22010386.

Charlton PH, Allen J, Bailón R, Baker S, Behar JA, Chen F, Clifford GD, Clifton DA, Davies HJ, Ding C, Ding X, Dunn J, Elgendi M, Ferdoushi M, Franklin D, Gil E, Hassan MF, Hernesniemi J, Hu X, Ji N, et al. (2023). The 2023 Wearable Photoplethysmography Roadmap. *Physiological Measurement* **44**, 111001. https://doi.org/10.1088/1361-6579/acead2.

Cohen J (1962). The Statistical Power of Abnormal-Social Psychological Research: A Review. *The Journal of Abnormal and Social Psychology* **65**, 145–153. https://doi.org/10.1037/h0045186.

Cronbach LJ, Meehl PE (1955). Construct Validity in Psychological Tests. *Psychological Bulletin* **52**, 281–302. https://doi.org/10.1037/h0040957.

Dawson ME, Schell AM, Filion DL (2007). The Electrodermal System. In: *Handbook of Psychophysiology*. Ed. by Gary Berntson, John T. Cacioppo, and Louis G. Tassinary. 3rd ed. Cambridge: Cambridge University Press, pp. 159–181. https://doi.org/10.1017/CBO9780511546396.

van Dongen N, van Bork R, Finnemann A, Haslbeck JMB, van der Maas HLJ, Robinaugh DJ, de Ron J, Sprenger J, Borsboom D (2025). Productive Explanation: A Framework for Evaluating Explanations in Psychological Science. *Psychological Review* **132**, 311–329. https://doi.org/10.1037/rev0000479.

Dyer EL, Richards BA (2025). Accepting “the bitter lesson” and embracing the brain’s complexity. The Transmitter. https://doi.org/10.53053/ORXM6480.

Eronen MI, Bringmann LF (2021). The Theory Crisis in Psychology: How to Move Forward. *Perspectives on Psychological Science* **16**, 779–788. https://doi.org/10.1177/1745691620970586.

Eum B, Dolbier S, Rangel A (2023). Peripheral Visual Information Halves Attentional Choice Biases. *Psychological Science* **34**, 984–998. https://doi.org/10.1177/09567976231184878.

Fernández-Alvarez J, Grassi M, Colombo D, Botella C, Cipresso P, Perna G, Riva G (2022). Efficacy of Bio- and Neurofeedback for Depression: A Meta-Analysis. *Psychological Medicine* **52**, 201–216. https://doi.org/10.1017/S0033291721004396.

Flake JK, Fried EI (2020). Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. *Advances in Methods and Practices in Psychological Science* **3**, 456–465. https://doi.org/10.1177/2515245920952393.

Frischkorn GT (2026). A shared vocabulary for reasoning about the laboratory as a whole. *Peer Community in Psychology*, 100132. https://doi.org/10.24072/pci.psych.100132.

Gärtner A, Leising D, Freyer N, Musfeld P, Lange J, Schönbrodt FD (2025). Responsible Research Assessment II: A Specific Proposal for Hiring and Promotion in Psychology. *Meta-Psychology* **9**. https://doi.org/10.15626/MP.2024.4604.

Gorgolewski KJ, Auer T, Calhoun VD, Craddock RC, Das S, Duff EP, Flandin G, Ghosh SS, Glatard T, Halchenko YO, Handwerker DA, Hanke M, Keator D, Li X, Michael Z, Maumet C, Nichols BN, Nichols TE, Pellman J, Poline JB, et al. (2016). The Brain Imaging Data Structure, a Format for Organizing and Describing Outputs of Neuroimaging Experiments. *Scientific Data* **3**, 160044. https://doi.org/10.1038/sdata.2016.44.

Goyal K, Borkholder DA, Day SW (2022). Dependence of Skin-Electrode Contact Impedance on Material and Skin Hydration. *Sensors* **22**, 8510. https://doi.org/10.3390/s22218510.

Hack L, Halperin I, Wolff W (2025). A Hitchhiker's Guide to Physical and Cognitive Effort. https://doi.org/10.31234/osf.io/gm7zf_v1.

Hallal PC, Pratt M (2020). Physical Activity: Moving from Words to Action. *The Lancet Global Health* **8**, e867–e868. https://doi.org/10.1016/S2214-109X(20)30256-4.

Hunt LT, Hayden BY (2017). A Distributed, Hierarchical and Recurrent Framework for Reward-Based Choice. *Nature Reviews Neuroscience* **18**, 172–182. https://doi.org/10.1038/nrn.2017.7.

Isbaner S, Báez-Mendoza R, Bothe R, Eiteljoerge S, Fischer A, Gail A, Gläscher J, Lüschen H, Möller S, Penke L, Priesemann V, Ruß J, Schacht A, Schneider F, Shahidi N, Treue S, Wibral M, Ziereis A, Fischer J, Kagan I, et al. (2025). Dyadic Interaction Platform: A Novel Tool to Study Transparent Social Interactions. *eLife* **14**, RP106757. https://doi.org/10.7554/eLife.106757.1.

Keil A, Debener S, Gratton G, Junghöfer M, Kappenman ES, Luck SJ, Luu P, Miller GA, Yee CM (2014). Committee Report: Publication Guidelines and Recommendations for Studies Using Electroencephalography and Magnetoencephalography. *Psychophysiology* **51**, 1–21. https://doi.org/10.1111/psyp.12147.

Khalili M, GholamHosseini H, Lowe A, Kuo MMY (2024). Motion Artifacts in Capacitive ECG Monitoring Systems: A Review of Existing Models and Reduction Techniques. *Medical & Biological Engineering & Computing* **62**, 3599–3622. https://doi.org/10.1007/s11517-024-03165-1.

Kothe C, Shirazi SY, Stenner T, Medine D, Boulay C, Grivich MI, Artoni F, Mullen T, Delorme A, Makeig S (2025). The Lab Streaming Layer for Synchronized Multimodal Recording. *Imaging Neuroscience* **3**, IMAG.a.136. https://doi.org/10.1162/IMAG.a.136.

Krakauer JW, Ghazanfar AA, Gomez-Marin A, MacIver MA, Poeppel D (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron* **93**, 480–490. https://doi.org/10.1016/j.neuron.2016.12.041.

Lonsdorf TB, Menz MM, Andreatta M, Fullana MA, Golkar A, Haaker J, Heitland I, Hermann A, Kuhn M, Kruse O, Meir Drexler S, Meulders A, Nees F, Pittig A, Richter J, Römer S, Shiban Y, Schmitz A, Straube B, Vervliet B, et al. (2017). Don't Fear 'Fear Conditioning': Methodological Considerations for the Design and Analysis of Studies on Human Fear Acquisition, Extinction, and Return of Fear. *Neuroscience & Biobehavioral Reviews* **77**, 247–285. https://doi.org/10.1016/j.neubiorev.2017.02.026.

Luck SJ (2014). *An Introduction to the Event-Related Potential Technique*. 2nd ed. Cambridge, MA, USA: MIT Press.

Machado A, Ferreira DF, Ferreira S, Almeida-Antunes N, Carvalho P, Melo P, Rocha N, Rodrigues MA (2025). A Systematic Review of Chest-Worn Sensors in Cardiac Assessment: Technologies, Advantages, and Limitations. *Sensors* **25**, 6049. https://doi.org/10.3390/s25196049.

Markiewicz CJ, Gorgolewski KJ, Feingold F, Blair R, Halchenko YO, Miller E, Hardcastle N, Wexler J, Esteban O, Goncavles M, Jwa A, Poldrack R (2021). The OpenNeuro Resource for Sharing of Neuroscience Data. *eLife* **10**. Ed. by Thorsten Kahnt, Chris I Baker, Nico Dosenbach, Michael J Hawrylycz, and Karel Svoboda, e71774. https://doi.org/10.7554/eLife.71774.

McLean S, Robertson S, Salmon P (2025). Complexity and Systems Thinking in Sport. *Journal of Sports Sciences* **43**, 1–5. https://doi.org/10.1080/02640414.2024.2388428.

Montagu JD, Coles EM (1966). Mechanism and Measurement of the Galvanic Skin Response. *Psychological Bulletin* **65**, 261–279. https://doi.org/10.1037/h0023204.

Mosley E, Laborde S (2024). A Scoping Review of Heart Rate Variability in Sport and Exercise Psychology. *International Review of Sport and Exercise Psychology* **17**, 773–847. https://doi.org/10.1080/1750984X.2022.2092884.

Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, Simonsohn U, Wagenmakers EJ, Ware JJ, Ioannidis JPA (2017). A Manifesto for Reproducible Science. *Nature Human Behaviour* **1**, 0021. https://doi.org/10.1038/s41562-016-0021.

Naughton M, Salmon PM, McLean S (2024). Where Do We Intervene to Optimize Sports Systems? Leverage Points the Way. *Journal of Sports Sciences* **42**, 566–573. https://doi.org/10.1080/02640414.2024.2352681.

Nebe S, Reutter M, Baker DH, Bölte J, Domes G, Gamer M, Gärtner A, Gießing C, Gurr C, Hilger K, Jawinski P, Kulke L, Lischke A, Markett S, Meier M, Merz CJ, Popov T, Puhlmann LM, Quintana DS, Schäfer T, et al. (2023). Enhancing Precision in Human Neuroscience. *eLife* **12**. Ed. by Chris I Baker, e85980. https://doi.org/10.7554/eLife.85980.

Noar SM, Zimmerman RS (2005). Health Behavior Theory and Cumulative Knowledge Regarding Health Behaviors: Are We Moving in the Right Direction? *Health Education Research* **20**, 275–290. https://doi.org/10.1093/her/cyg113.

Nyström M, Hooge ITC, Hessels RS, Andersson R, Hansen DW, Johansson R, Niehorster DC (2025). The Fundamentals of Eye Tracking Part 3: How to Choose an Eye Tracker. *Behavior Research Methods* **57**, 67. https://doi.org/10.3758/s13428-024-02587-x.

Pearl J (2009). *Causality*. 2nd ed. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511803161.

Pernet CR, Appelhoff S, Gorgolewski KJ, Flandin G, Phillips C, Delorme A, Oostenveld R (2019). EEG-BIDS, an Extension to the Brain Imaging Data Structure for Electroencephalography. *Scientific Data* **6**, 103. https://doi.org/10.1038/s41597-019-0104-8.

Pessoa L (2008). On the Relationship between Emotion and Cognition. *Nature Reviews Neuroscience* **9**, 148–158. https://doi.org/10.1038/nrn2317.

Pessoa L (2019). Embracing Integration and Complexity: Placing Emotion within a Science of Brain and Behaviour. *Cognition and Emotion* **33**, 55–60. https://doi.org/10.1080/02699931.2018.1520079.

Pessoa L (2022). *The Entangled Brain: How Perception, Cognition, and Emotion Are Woven Together*. The MIT Press. https://doi.org/10.7551/mitpress/14636.001.0001.

Pinti P, Scholkmann F, Hamilton A, Burgess P, Tachtsidis I (2019). Current Status and Issues Regarding Preprocessing of fNIRS Neuroimaging Data: An Investigation of Diverse Signal Filtering Methods Within a General Linear Model Framework. *Frontiers in Human Neuroscience* **12**. https://doi.org/10.3389/fnhum.2018.00505.

Piray P (2026). Addressing Low Statistical Power in Computational Modelling Studies in Psychology and Neuroscience. *Nature Human Behaviour* **10**, 347–356. https://doi.org/10.1038/s41562-025-02348-6.

Scharfen HE, Memmert D (2024). The Model of the Brain as a Complex System: Interactions of Physical, Neural and Mental States with Neurocognitive Functions. *Consciousness and Cognition* **122**, 103700. https://doi.org/10.1016/j.concog.2024.103700.

Schönbrodt FD, Gärtner A, Frank M, Gollwitzer M, Ihle M, Mischkowski D, Phan LV, Schmitt M, Scheel AM, Schubert AL, Steinberg U, Leising D (2025). Responsible Research Assessment I: Implementing DORA and CoARA for Hiring and Promotion in Psychology. *Meta-Psychology* **9**. https://doi.org/10.15626/MP.2024.4601.

Schüler J, Heino MTJ, Balagué N, Chater AM, Gruber M, Kanning M, Keim D, Mier D, Moreno-Villanueva M, Nussbeck FW, Pruessner J, Shafie T, Schwenk M, Bieleke M (2025). A Complex Systems View on Physical Activity with Actionable Insights for Behaviour Change. *Nature Human Behaviour* **9**, 1793–1801. https://doi.org/10.1038/s41562-025-02279-2.

Schultz T, Putze F, Reisenhofer R, Fehr T, Meier M, Mason C, Ahrens F (2024). LabLinking: Theory, Framework, and Solutions of Connecting Laboratories for Distributed Human Experiments. *Discover Applied Sciences* **6**, 448. https://doi.org/10.1007/s42452-024-06122-7.

Shamay-Tsoory SG, Mendelsohn A (2019). Real-Life Neuroscience: An Ecological Approach to Brain and Behavior Research. *Perspectives on Psychological Science* **14**, 841–859. https://doi.org/10.1177/1745691619856350.

Strain T, Flaxman S, Guthold R, Semenova E, Cowan M, Riley LM, Bull FC, Stevens GA, Raheem RA, Agoudavi K, Anderssen SA, Alkhatib W, Aly EAH, Anjana RM, Bauman A, Bovet P, Moniz TB, Bulotaite˙ G, Caixeta R, Monteiro EC, et al. (2024). National, Regional, and Global Trends in Insufficient Physical Activity among Adults from 2000 to 2022: A Pooled Analysis of 507

Population-Based Surveys with 5·7 Million Participants. *The Lancet Global Health* **12**, e1232–e1243. https://doi.org/10.1016/S2214-109X(24)00150-5.

Tachtsidis I, Scholkmann F (2016). False Positives and False Negatives in Functional Near-Infrared Spectroscopy: Issues, Challenges, and the Way Forward. *Neurophotonics* **3**, 031405. https://doi.org/10.1117/1.NPh.3.3.031405.

Tremblay S, Testard C, Inchauspé J, Petrides M (2023). Epiphenomenal Neural Activity in the Primate Cortex. bioRxiv. https://doi.org/10.1101/2022.09.12.506984.

Vul E, Harris C, Winkielman P, Pashler H (2009). Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition1. *Perspectives on Psychological Science* **4**, 274–290. https://doi.org/10.1111/j.1745-6924.2009.01125.x.

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, et al. (2016). The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* **3**, 160018. https://doi.org/10.1038/sdata.2016.18.

Wolff W, Hack L, Halperin I, Holgado D, Englert C, Zech A, Schüler J, Bieleke M, Shenhav A, Martarelli C (2025).When Effort Fails: Instances and Reasons for Effort-Performance Decoupling. PsyArXiv. https://doi.org/10.31234/osf.io/3uces_v1.

Wu H, Yang G, Zhu K, Liu S, Guo W, Jiang Z, Li Z (2021). Materials, Devices, and Systems of On-Skin Electrodes for Electrophysiological Monitoring and Human–Machine Interfaces. *Advanced Science* **8**, 2001938. https://doi.org/10.1002/advs.202001938.

Xiang MQ, Hou XH, Liao BG, Liao JW, Hu M (2018). The Effect of Neurofeedback Training for Sport Performance in Athletes: A Meta-Analysis. *Psychology of Sport and Exercise* **36**, 114–122. https://doi.org/10.1016/j.psychsport.2018.02.004.

Yücel MA, v Lühmann A, Scholkmann F, Gervain J, Dan I, Ayaz H, Boas D, Cooper RJ, Culver J, Elwell CE, Eggebrecht A, Franceschini MA, Grova C, Homae F, Lesage F, Obrig H, Tachtsidis I, Tak S, Tong Y, Torricelli A, et al. (2021). Best Practices for fNIRS Publications. *Neurophotonics* **8**, 012101. https://doi.org/10.1117/1.NPh.8.1.012101.