D I S S E R T A T I O N

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

Researching Neural Correlates of Human Decisions from Experience using Electroencephalography

vorgelegt von

Stefan Appelhoff, M.Sc.

am Fachbereich Erziehungswissenschaft und Psychologie

der Freien Universität Berlin

Berlin, 2022

Erstgutachter: Prof. Dr. Ralph Hertwig
Zweitgutachter: Prof. Dr. Radoslaw Martin Cichy

Tag der Disputation: Mittwoch, 06. Juli 2022

Acknowledgements

This dissertation and the work that went into it would not have been possible without the help and support I received from mentors, colleagues, friends, and family.

First of all I want to thank Bernhard Spitzer. Thank you, Bernhard, for being my mentor throughout this journey. I could always (and at literally any time) approach you with any problems I had – be it scientific or otherwise. You took the time to think through tricky problems together with me, challenged me to try new analyses and look into different literature, and nudged me to continue on my path whenever I stopped.

I also want to thank Ralph Hertwig for many things. Thanks for giving me the opportunity to do my PhD in such an active and interdisciplinary place like ARC at the MPIB. I deeply appreciate that you allowed and encouraged me to pursue projects like the Brain Imaging Data Structure, or the Google Summer of Code in 2019. Thank you also for reminding me regularly to not drift off too much into that direction, whenever time on my PhD started to run short.

Next, my colleagues at ARC at the MPIB deserve my thanks and appreciation. Starting from my first day at ARC, I enjoyed your company and the welcoming culture at place. I still remember Christina Leuker asking me whether I'd like to join for lunch on that day. Many people have left and many new people have joined ARC since I began my journey in April 2018, but the generally open and supportive culture has not changed, not even due to the Corona pandemic. I am thankful when thinking back to our retreats, to our jam sessions, to our restaurant and bar visits, and also – more professionally – to all the times I received feedback or help on projects (including help from the IT and administration at ARC and the MPIB in general!). Thank you Anastasia Kozyreva and Linda Kerbl for organizing the Summer Institute on Bounded Rationality together with me in 2019 – you made that challenging task fun (mostly, because it was still an awful amount of work). Thank you Thorsten Pachur for having helpful and encouraging words for me at our meetings. For this dissertation, I had to collect a lot of data (and some of it during a pandemic). I want to thank all those who lent a helping hand (you are mentioned explicitly in the chapters). Thank you, all members of CNARC ("cognitive neuroscience at ARC") for just being a lovely crowd. I also want to thank Dirk Ostwald. Thank you, Dirk, for supporting me during my Master's degree and at the beginning of my PhD, and for being an inspiration with regards to *thorough* scientific work. Lastly, before turning to support I received from my personal background, I want to thank the open source software development and open science communities centered around the MNE-Python and BIDS organizations – I continue to learn a lot from many of you, and I am very thankful to be a part of this.

It goes without saying that there are times in which work is just hard, and colleagues and mentors are not the only support that is needed to complete a PhD. I want to thank my friends for having my back. I am thankful for all the things we do, be it visiting each other, generally staying in

touch, playing board games, tennis, or volleyball, be it hiking in Norway or Portugal, or playing Dungeons and Dragons over video chat while nobody wanted to go outside (again, the pandemic).

My family supported me throughout this (at times very rough) ride, and I deeply appreciate that. You always believed that I can do it, and that was a source of strength for me.

Thank you, Lorna, for always being there for me.

Summary

People routinely collect information about choice alternatives before deciding between them. For example before buying a new car, a customer may test drive several alternatives before making a choice. Such decisions from experience (DfE) are characterized by active or passive (i.e., through observation) sampling processes that may involve categorizing and evaluating each sample, and finally integrating all available samples into a choice. After a short introduction (chapter 1), in part one of this dissertation I present two studies from the domain of decisions from experience, addressing separate but closely related questions. The first study (chapter 2) deals with the question of how control over the sampling process impacts on the final choice. In how far do full, partial, and no control over sampling differ in terms of their neurocognitive processes, and are some levels of control more beneficial than others? The second study (chapter 3) focuses on processes related to evaluating individual samples in a sequence. How are individual samples weighted before they are integrated into a choice? For both studies, I employ Electroencephalography (EEG) recordings, relating behavior and brain measures. In part two of this dissertation, I then shift from decisions from experience to more methodological issues related to EEG research and present two further studies. The third study of this dissertation (chapter 4) addresses an emerging hardware problem: The parallel port, the previous gold standard for sending event markers, is slowly disappearing from commercially available computers. What are potential alternatives, how reliable and performant are they, and how can they be built and operated on a budget? In the fourth study (chapter 5), my co-authors and I present an extension to an existing and commonly used data standard (BIDS) for EEG data, with the aim of improving the foundations of good research data management and reproducibility of scientific results. Finally, I conclude with a chapter (chapter 6) to briefly summarize the other chapters and delineate open questions and future directions.

Zusammenfassung

Menschen sammeln routinemäßig Informationen über Alternativen, bevor sie sich zwischen ihnen entscheiden. Beispielsweise kann ein Kunde vor dem Kauf eines neuen Autos mehrere Alternativen testen, bevor er eine Entscheidung trifft. Solche Verfahren sind durch aktive oder passive (d.h. durch Beobachtung) Stichprobenprozesse gekennzeichnet, die eine Kategorisierung und Bewertung der einzelnen Proben und schließlich die Integration aller verfügbaren Proben in eine Entscheidung umfassen können. Nach einer kurzen Einführung (chapter 1), stelle ich im ersten Teil dieser Dissertation zwei Studien aus dem Bereich "Entscheidungen nach Erfahrung" vor, die sich mit unterschiedlichen, aber eng verwandten Fragen befassen. Die erste Studie (chapter 2) beschäftigt sich mit der Frage, wie sich die Kontrolle über den Stichprobenprozess auf die endgültige Entscheidung auswirkt. Inwieweit unterscheiden sich vollständige, teilweise und keine Kontrolle über die Probenahme in Bezug auf ihre neurokognitiven Prozesse, und sind einige Stufen der Kontrolle vorteilhafter als andere? Die zweite Studie (chapter 3) konzentriert sich auf Prozesse im Zusammenhang mit der Bewertung einzelner Proben in einer Sequenz von Proben. Wie werden die einzelnen Proben gewichtet bevor sie in eine Entscheidung integriert werden? Für beide Studien verwende ich Elektroenzephalographie (EEG)-Aufnahmen, und setze Verhaltens- und Gehirnmessungen in Beziehung. Im zweiten Teil dieser Dissertation wende ich mich dann von "Entscheidungen nach Erfahrung" zu eher methodischen Fragen im Zusammenhang mit EEG Forschung und stelle zwei weitere Studien vor. Die dritte Studie dieser Dissertation (chapter 4) befasst sich mit einem neuen Hardwareproblem: Die parallele Schnittstelle, der bisherige Goldstandard zum Senden von Ereignismarkern, verschwindet langsam von handelsüblich erwerbbaren Computern. Was sind mögliche Alternativen, wie zuverlässig und leistungsfähig sind sie, und wie können sie mit einem geringen Budget gebaut und betrieben werden? In der vierten Studie (chapter 5) stellen meine Mitautoren und ich eine Erweiterung eines bestehenden und häufig verwendeten Datenstandards (BIDS) für EEG-Daten vor. Dabei verfolgen wir das Ziel, die Grundlagen für ein gutes Forschungsdatenmanagement und die Reproduzierbarkeit wissenschaftlicher Ergebnisse zu verbessern. Zum Abschluss stelle ich ein Kapitel vor (chapter 6), das die anderen Kapitel kurz zusammenfasst und offene Fragen und zukünftige Richtungen beschreibt.

Contents

Acronyms

BEP BIDS Extension Proposal. 6, 7, 72, 79

BIC Bayesian Information Criterion. 49

BIDS Brain Imaging Data Structure. vii, ix, 6, 7, 18, 30, 67–73, 77, 79, 80

BNT Berlin Numeracy Test. 18

COBIDAS Committee on Best Practice in Data Analysis and Sharing. 72

COM Communication Port. 57, 58

CPP centro-parietal positivity. 22–24, 77

CRediT Contributor Roles Taxonomy. 87

DfD decisions from description. 1, 2

DfE decisions from experience. vii, 1, 2, 5

DIY Do-it-yourself. 63

ECG electrocardiogram. 18, 19, 48, 49

EDF European Data Format. 70, 71

EEG Electroencephalography. vii, ix, 1, 5, 6, 15, 18, 19, 21–23, 26, 28, 29, 37–40, 48, 49, 63, 67–72, 78–80, 94

EOG electrooculogram. 18, 19, 48, 49

ERP Event Related Potential. 19, 20, 22–27, 40, 42–44, 49–51, 58, 78

FAIR Findable, Accessible, Interoperable, Reusable. 68, 72

FDR False Discovery Rate. 27, 42, 43

fMRI Functional Magnetic Resonance Imaging. 68, 71

HED Hierarchical Event Descriptor. 70

ICA independent component analysis. 19, 49

iEEG Intracranial electroencephalography. 6, 72

LSL Lab Streaming Layer. 60–62, 79

MCU Microcontroller Unit. 1, 6, 57–64, 77–79

MEG Magnetoencephalography. 6, 68

MoBI Mobile Brain/Body Imaging. 79

MRI Magnetic Resonance Imaging. 6, 68–70, 72, 80

NODDI Neurite orientation and dispersion density imaging. 71

OHBM Organization for Human Brain Mapping. 72

PCI Peripheral Component Interconnect. 58

PET Positron Emission Tomography. 6, 68

RDM representational dissimilarity matarix. 20, 21, 24–26, 41, 50, 51

RSA representational similarity analysis. 20, 23–25, 37, 38, 40–44, 50, 51, 78

SD Standard Deviation. 17

SE Standard Error. 16, 21, 23, 25, 27, 39–41, 44

SPI Serial Peripheral Interface. 59

TTL Transistor-transistor logic. 57–61, 63, 79

USB Universal Serial Bus. 5, 18, 57–59, 61, 64

1 | General Introduction

Imagine it is Friday evening after a long and grueling work week and you just entered a wine shop that you have wanted to visit for a long time. Naturally, you want to acquire a good bottle of wine to enjoy over the weekend, but it must not be too expensive lest you spent all your hard earned money in that single purchase. To get a first impression, you may start with browsing the catalogue of available wines to arrive at a decision whether this wine shop is more or less pricey than the other wine shops you know. Having your tight budget in mind, you might even try to estimate whether on average the white wines or the red wines in this shop are more expensive. Next, you may request to try a few samples, to not only get the most affordable wine (judging from the catalogue), but also the best tasting one, before you finally decide for a bottle to buy.

All of these three actions our imaginary agent may undertake have something in common: They lead to decisions from experience, that is, decisions based on sequential samples that need to be categorized, integrated, and formed into a choice. In the first part of this dissertation, I will present two empirical studies covering the neural correlates of decisions from experience, as revealed through Electroencephalography (EEG).

The second part of this dissertation has a focus on practical, methodological matters directly related to empirical EEG research. I will propose to record experimental event markers with a Microcontroller Unit (MCU) based solution that can serve as a viable alternative to the previous gold standard that is rapidly disappearing from commercially available equipment. Furthermore, I will describe recent advances with regards to standardized documentation and sharing of research data, in order to improve reproducibility of scientific results.

In the following I will briefly outline and contextualize each of these two parts.

1.1 Decisions from Experience

Since the early 2000s, the term decisions from experience (DfE) has been frequently used in behavioral research to describe contexts where study participants initially have little or no knowledge about a number of payoff distributions, and are then instructed to sample from the options to maximize either their cumulative payoff or their prospective payoff in a single choice after the sampling phase (Hertwig, 2015; Hertwig et al., 2004; Weber et al., 2004; Wulff et al., 2018). DfE as a term is most meaningful when seen as one endpoint on a continuum with decisions from description (DfD) as the opposite endpoint. In contrast to DfE, DfD happen whenever an agent receives a full description of all available choice options, and the probabilities associated with each of their outcomes (e.g., Gigerenzer et al., 2005). For example when deciding whether or not to take a drug based on statistical reports of adverse effects on the drug's package insert. Contrasting DfE and DfD in human decision-makers has revealed the so-called description-experience-gap: Humans

tend to underweight rare events in decisions from experience, but overweight them in decisions from description (Hertwig & Erev, 2009; Rakow & Newell, 2010).

However, decisions from experience are a broad and relevant research topic in their own right, with a long standing tradition aside from its contrast to decisions from description (e.g., Anderson, 1964; Busemeyer, 1982, 1985; Edwards, 1961; Robbins, 1952). In research areas outside of behavioral sciences such as in cognitive neuroscience or computer science, DfE is not often used as a term, perhaps because its natural counterpoint DfD receives a much smaller focus compared to in behavioral research. Instead, DfE are being researched more generally under terms such as evidence accumulation, sequential sampling, or sequential integration (Cheadle et al., 2014; Kang & Spitzer, 2021; Shadlen & Shohamy, 2016; Spitzer et al., 2017; Tsetsos et al., 2012; Wyart et al., 2012; Wyart et al., 2015), or under the umbrella of reinforcement learning and exploration and exploitation (Sutton & Barto, 2018), in form of the multi-armed bandit paradigm and its many derivatives such as the fully explorative bandit or best-arm identification problem (Audibert & Bubeck, 2010; Bubeck et al., 2009; Daw et al., 2006; Even-Dar et al., 2006; Gabillon et al., 2012; Shahrampour et al., 2017).

In this dissertation, decisions from experience will refer to all processes that involve actively or passively (i.e., through observation) sampling outcomes from one or several distributions, and forming the sampled information into a final choice that can be correct or incorrect with regards to some aspects of the sampled distributions.

Let us now come back to our wine shop example, and see how the imaginary agent's decisions from experience map onto the laboratory tasks that I employ in this dissertation. First, our agent tried to browse the catalogue to get an impression of how pricey the shop was. This corresponds to the *averaging task* (Figure 1.1, *left*), where an agent observes a rapid sequence of Arabic digits from 1 to 9 and is then asked whether the arithmetic mean of the sequence was higher or lower than 5. Second, our agent made an (arguably futile) attempt to get the best deal by setting white wines against red wines, trying to see whether one category was more expensive than the other. This corresponds to the *comparison task* (Figure 1.1, *middle*), where an agent observes digits as in the averaging task, but each digit is additionally marked with a color and the agent is asked whether the arithmetic mean of digits in one color was higher than the other. Third and finally, our agent took the opportunity to taste several wines and purchase a single bottle. This corresponds to the *sampling paradigm* (Figure 1.1, *right*), where an agent can actively sample from one of several unknown payoff distributions until they decide to stop sampling and choose one of the options to receive a payoff in the form of a random draw from that option.

I will now briefly outline some of the key research questions that have been addressed with these tasks, and specify the open questions in the current literature that are addressed with this dissertation.

1.1.1 The sampling paradigm: human exploration and search

Unlike many other paradigms of decisions from experience, the sampling paradigm strictly separates exploration and exploitation. That is, the typical trade-off that an agent faces of whether to use their limited time and resources to *explore* an unknown but potentially rewarding option or whether to *exploit* the option that is currently thought to be the most rewarding, is suspended (Sutton & Barto, 2018). Instead, study participants first freely (and at no expense) explore all available choice options and are given the chance to learn about their underlying payoff distributions (exploration). Only once a participant willingly terminates their exploration, the exploitation

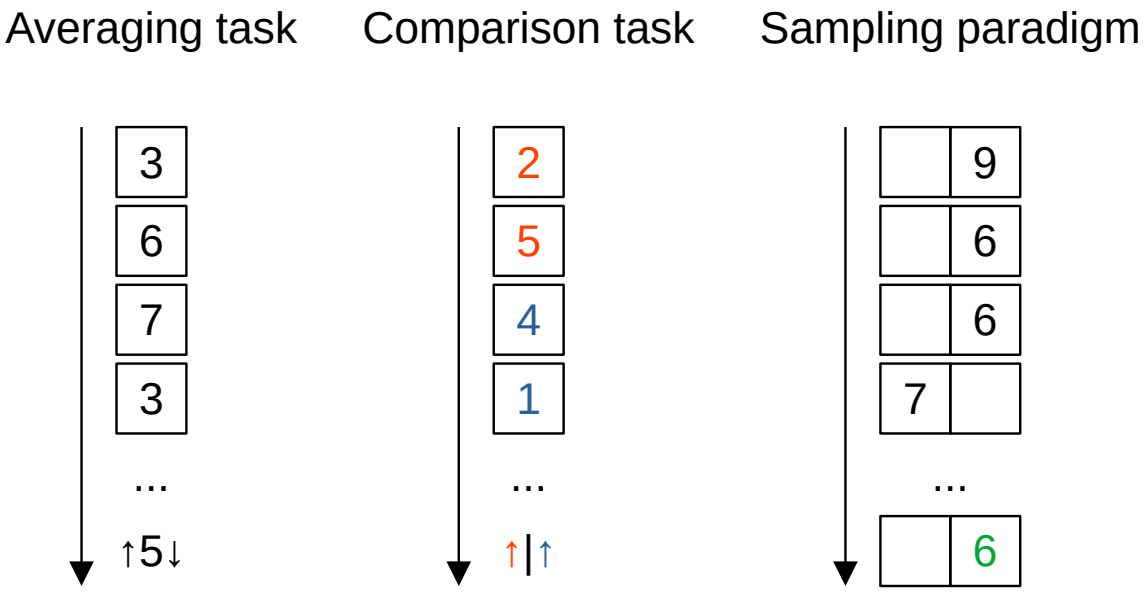

Figure 1.1. Laboratory tasks employed in this dissertation. Left: In the averaging task, participants observe a rapid stream of 10 digits between 1 and 9 at a rate of about 3 digits per second. After having seen all digits, participants need to decide whether the arithmetic mean of the sequence was above or below 5. **Middle:** In the comparison task, participants observe a rapid stream of digits as in the averaging task (*left*), however each digit sample has one of two meaningful color categories (here: red and blue). After having seen all digits, participants need to decide whether the red digits or the blue digits had a higher mean. **Right:** In the sampling paradigm, participants can actively sample two previously unknown payoff distributions (here: left and right "squares") for as long as they wish. When the participant stops sampling, they make a final choice between the two payoff distributions and a random draw from that choice is awarded to the participant as a bonus (indicated with green color).

phase begins, that is: The participant would select one of the available choice option to receive a random draw from it as their reward.

This feature allows a clear view on self-directed information search, without having to first disentangle whether an agent's action was explorative or exploitative in a given context (see e.g., Daw et al., 2006). Thus, the sampling paradigm has been used extensively to address questions such as (i) "Which factors determine the extend of exploration (number of drawn samples) in self-directed search?" (Lejarraga et al., 2012; Rakow et al., 2008; Spaniol & Wegier, 2012), (ii) "What are predictors of stopping to search?" (Hertwig & Pleskac, 2010; Markant et al., 2015; Ostwald et al., 2015), or (iii) "Do human agents resort to stereotypical search patterns?" (Hills & Hertwig, 2010).

In much of the research employing the sampling paradigm, authors have furthermore controlled the *extent* to which information search is self-directed, for example by limiting the overall amount of samples that can be drawn from payoff distributions (Gonzalez & Mehlhorn, 2016; Hau et al., 2008), or by suspending the search altogether and letting the computer draw samples for the agent (Rakow et al., 2008). Such manipulations either served as a contrast to "fully free" self-directed information search, or to bias self-directed information search in directions that would make the subsequent choices more comparable to those in decisions from description (Rakow & Newell, 2010). Yet, to date few studies have directly compared how different degrees of control over self-directed sampling and exploration affect subsequent choice.

In chapter 2, I will present a study in which my co-authors and I show that full control over sampling, compared to partial control and no control, leads to a distinct benefit in behavioral accuracy. Furthermore, through EEG measures and multivariate analysis thereof, we shed new light on the neurocognitive processes that underlie these behavioral benefits.

1.1.2 "Averaging" and "comparison" tasks: approximate numerical judgment and non-linear weighting of samples

When forming decisions based on numerical samples, not all samples are judged equal. Imagine that you see ten digits between 1 and 9 and need to decide whether their average was larger than 5 (this is the averaging task, see Figure 1.1, *left*). A reasonable policy would be to give a sample "6" in the sequence just as much weight towards the final decision as a sample "4". Similarly, a sample "7" should receive just as much more (absolute) weight compared to a "6", as a "3" would receive compared to a "4". In many contexts however, decision-makers do not employ such "linear" weighting policies. Rather, they are often found to either overweight inlying values (e.g., 4 and 6 in our example) in terms of decision weight relative to outlying values (e.g., 1 and 9 in our example), or to exhibit the opposite tendency of overweighting outlying values. Relative overweighting of inlying values has been called "compression", and is commonly observed from psychophysics to economic choice (Bernoulli, 1954; Fechner, 1860; Juechems et al., 2021; Tversky & Kahneman, 1992). The opposite tendency, "anti-compression", has also been observed – most often in decisions from experience – both in humans (e.g., Spitzer et al., 2017; Tsetsos et al., 2012), and in non-human animals (Ludvig et al., 2014; Madan et al., 2019). In addition to compression or anti-compression, decision-makers may show a general bias towards either large or small values. These concepts are visualized and further described in Figure 1.2.

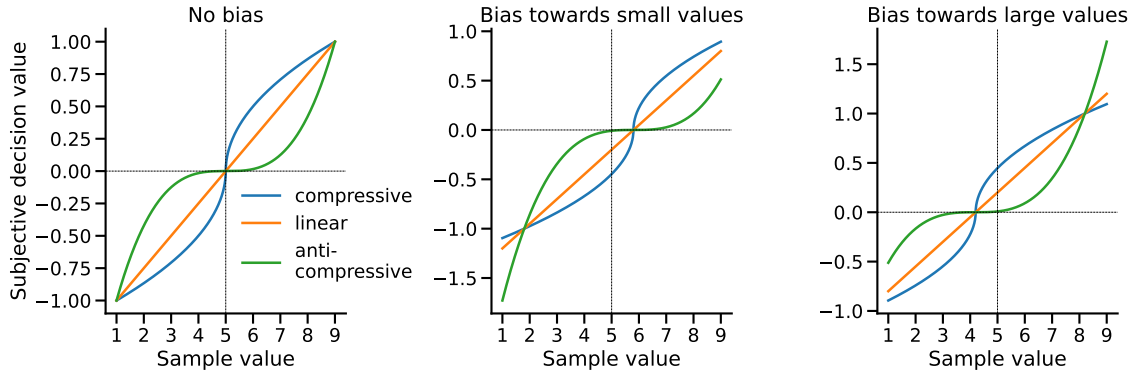

Figure 1.2. Schematic depiction of linear and non-linear weighting policies. Potential sample values (here: 1 to 9 in a sequence of random digits) are mapped onto subjective decision values (arbitrary units), which indicate how much sway a given sample holds over the final choice (here: whether the mean of a sequence of digits higher than 5). **Left:** During compressive weighting (blue line), inlying values are overweighted, that is, the highest rate of change in the function is at the inlying values. Such a weighting policy gives rise to a "diminishing sensitivity" to larger values in economic choice (Tversky & Kahneman, 1992). The opposite can be seen during anti-compressive weighting (green line), where outlying values are overweighted, and the highest rate of change in the function is at outlying values. **Middle/Right:** Same conventions as in *left*, but with an additional bias towards small/large values.

Weighting of sample values has major consequences for subsequent choice, and arguably the sampling process itself when considering the impact that, for example, the history of previous choices has on subsequent choices (e.g., Urai et al., 2019). Moreover, a large range of seemingly irrational human behaviors can be explained by alternative weighting policies and recast as "optimal behavior under constraints", "bounded rationality", "resource rationality", or "rational inattention" (Gigerenzer & Brighton, 2009; Lieder & Griffiths, 2020; Sims, 2003; Summerfield & Tsetsos, 2015). But why do decision-makers employ compression in some contexts, and anti-compression in others? Recent findings suggest that, while compression is far more commonly observed, anti-compression can arise in capacity-limited observers who must focus their resources on a subset of samples at the expense of others (Clarmann von Clarenau et al., 2022). However – as the authors themselves

discuss – their framework of explaining when compression or anti-compression arise does not lend itself to *predicting* which contexts would promote either behavior. For such predictions, neuroscientific approaches to uncover mechanisms of over- and underweighting, as well as to determine resources that particular tasks require, are indispensable.

In chapter 3, I will present a study in which my co-authors and I employ two different task instructions for the same physical stimuli that lead to compression and anti-compression in behavioral choice. Intriguingly, we find that the corresponding neural geometries indicated anti-compression in both cases. This suggests enhanced processing of extreme (outlying) values as a "default" in the brain, and raises the question at which point the eventual downweighting process occurs that manifests itself in the compression observed in behavior.

1.2 Improving two aspects of electroencephalography research

For the second part of this dissertation, we take a step back from the empirical questions on decisions from experience, and how they can be researched using EEG. Instead, we will deal with two concrete problems that arise with EEG research during the scientific process.[1] Namely (i) how to properly mark events in the EEG data stream for later analysis, and (ii) how to document and organize neuroimaging data to maximize reproducibility.

1.2.1 The parallel port: replacements for the disappearing gold standard

If you are reading this dissertation and were born after the year 2000, chances are high that you do not know the "parallel port". Other readers may remember the parallel port as something on your PC to connect a printer device to. However regardless of age and knowledge about computer devices, almost all readers who have performed EEG research will know about the parallel port, and a major practical problem associated with it.

The parallel port has traditionally been used as an interface between computers and EEG measurement equipment to synchronize the exact time when a study participant experiences a certain event (e.g., seeing a stimulus presented on screen) with the corresponding voltage measurement in micro Volts on their scalp (EEG). The reason for its adoption in research equipment were two-fold: (i) through use of direct current signals that are directly translated into digital states, the parallel port can achieve microsecond resolution of data transmission, and (ii) the parallel port was commonly available on consumer-grade personal computers. While reason (i) is still valid and a good reason to keep using the parallel port, reason (ii) reflects the major issue I alluded to above. In brief: the parallel port has been largely replaced by the USB port, and researchers need to find replacements to interface their (potentially slightly older but still robust) research hardware with modern computer hardware.

In chapter 4, I present a study in which my co-author and I test various alternatives to the dated gold standard and discuss why transparent testing, reporting, and guidelines of such solutions can help to improve EEG research (see also Figure 1.4).

[1] In fact, both problems came up for me personally during my work on decisions from experience and I wanted to solve them along the way.

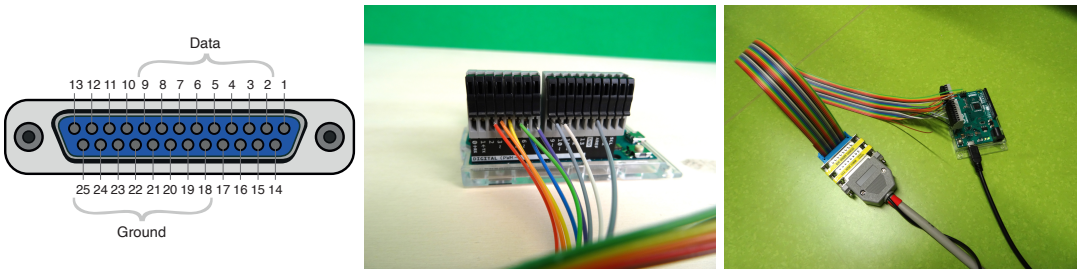

Figure 1.3. The parallel port and an adapter. Left: Schematic of the parallel port connector with 25 pins. For our purposes, only 8 data pins (corresponding to 1 byte of data) and any one of the 8 ground pins are relevant (see labels in the schematic). **Middle:** 8 data and 1 ground cable connecting to a Microcontroller Unit (MCU) that served as an adapter from USB to parallel port (here, the MCU is an "Arduino Leonardo"). **Right:** USB cable from a computer (black) connecting to the MCU, and then to the EEG amplifier via a parallel port cable (gray)

1.2.2 BIDS: a community-developed standard for organizing and documenting data

Conducting a very simple psychological study might result in a dataset that can easily be fit into a single `csv` file, with each row representing a participant, and with an accompanying `README` file to describe the columns and a few further details. However, even for this simplistic example[2], some problems are apparent:

- Which variables are recorded during the study, and are all present in the data file?
- How are the variables documented in the `README` file?
- To what extent is metadata about the experiment procedure reported in the `README` file?

Questions such as these pile up in more complicated experimental designs. For example, studies involving EEG recordings will additionally contain the EEG data, which may be in any vendor-specific file format that might be difficult to read with the software at hand. This dataset might furthermore contain anatomical MRI brain scans for source-reconstruction of the EEG data, and so on. More generally, idiosyncratic ways to organize data can easily lead to misunderstandings between collaborators, missing (meta-)data, and a waste of time to re-organize data to fit the structure and metadata expected by analysis scripts (or vice versa). All of these negatively impact the reproducibility of scientific research, often even in the same lab, and potentially by the same person at different points in time.

The Brain Imaging Data Structure (BIDS) is an emerging standard for dataset organization and documentation to address these problems (https://bids.neuroimaging.io/). With its inception in 2015, BIDS now covers rules and guidelines for multiple neuroimaging modalities such as MRI, MEG, EEG, iEEG, and PET (Gorgolewski et al., 2016; Holdgraf et al., 2018; Niso et al., 2018; Norgaard et al., 2022; Pernet et al., 2019). BIDS is a community-built and maintained, and an evolving standard. That is, all changes and additions to the standard are proposed and executed via transparent and inclusive processes – such as a BIDS Extension Proposal (BEP) – that allow each contributor to have a say (see Figure 1.2 for an overview of the community structure).

In chapter 5, I present the extension of BIDS to include rules and guidelines for EEG data, which my co-authors and I merged into the BIDS specification in 2019.

[2]The simplistic example is also an optimistic example, because we assume the data is shared in a commonly readable data format, `csv`, and a `README` file is supplied. It is not unlikely that instead, just a single proprietary data format file without enough documentation is shared. For example a single `SPSS` file.

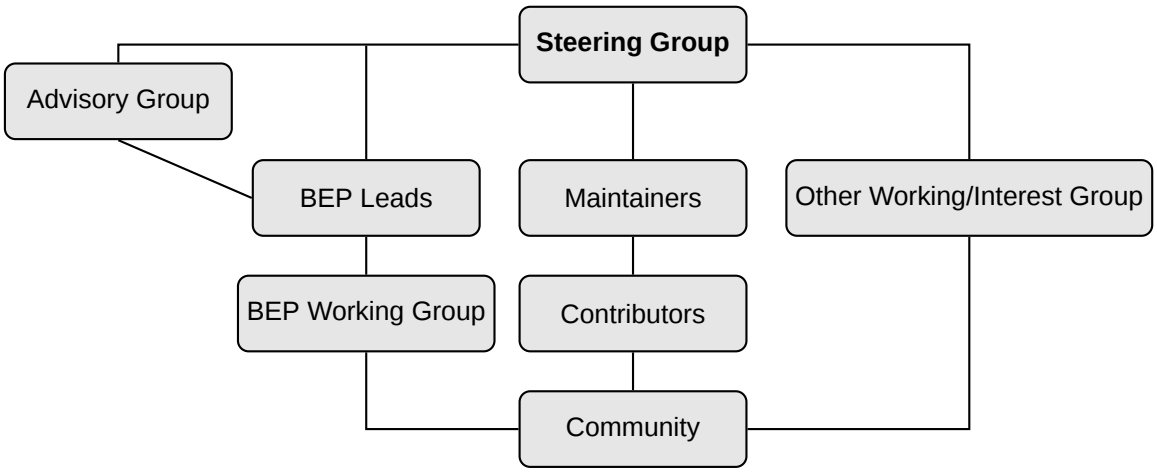

Figure 1.4. Schematic of the BIDS community structure. The **community** broadly consists of all individuals who used or are generally interested in BIDS. Any community member may become a contributor, for example by fixing typos in the specification text, or by pointing out and fixing inconsistencies. **Maintainers** are very active community members who support the general infrastructure, help contributors with their plans, and develop long-term plans for the project. The maintainers work closely with the **steering group**, an elected group of individuals who make decisions pertaining to the longevity and sustainability of the project, and who hold authority of last resort in addressing conflicts among groups and alleged violations of the Code of Conduct. When there is a need for substantial changes or additions to the BIDS specification, the community may for a **BIDS Extension Proposal (BEP) working group**, which is headed by **BEP leads**. Past BEP leads become part of the **advisory group**, who stay in close contact with current maintainers and the steering group. Any **other working or interest group** may be formed by the community, subject to steering group approval. For more information, see the BIDS Governance Document (https://bids.neuroimaging.io/governance.html).

1.3 References

Anderson, N. H. (1964). Test of a model for number-averaging behavior [Place: US Publisher: Psychonomic Society]. *Psychonomic Science*, *1*(7), 191–192. https://doi.org/10.3758/BF03342858

Audibert, J.-Y., & Bubeck, S. (2010). Best Arm Identification in Multi-Armed Bandits, 13 p. Retrieved February 17, 2022, from https://hal-enpc.archives-ouvertes.fr/hal-00654404

Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica*, *22*(1), 23. https://doi.org/10.2307/1909829

Bubeck, S., Munos, R., & Stoltz, G. (2009). Pure Exploration in Multi-armed Bandits Problems. In R. Gavaldà, G. Lugosi, T. Zeugmann, & S. Zilles (Eds.), *Algorithmic Learning Theory* (pp. 23–37). Springer. https://doi.org/10.1007/978-3-642-04414-4_7

Busemeyer, J. R. (1982). Choice behavior in a sequential decision-making task. *Organizational Behavior and Human Performance*, *29*(2), 175–207. https://doi.org/10.1016/0030-5073(82)90255-0

Busemeyer, J. R. (1985). Decision making under uncertainty: A comparison of simple scalability, fixed-sample, and sequential-sampling models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(3), 538–564. https://doi.org/10.1037/0278-7393.11.3.538

Cheadle, S., Wyart, V., Tsetsos, K., Myers, N., de Gardelle, V., Herce Castañón, S., & Summerfield, C. (2014). Adaptive Gain Control during Human Perceptual Choice. *Neuron*, *81*(6), 1429–1441. https://doi.org/10.1016/j.neuron.2014.01.020

Clarmann von Clarenau, V., Pachur, T., & Spitzer, B. (2022). *Over- and Underweighting of Extreme Values in Decisions from Sequential Samples* (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/6yj4r

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879. https://doi.org/10.1038/nature04766

Edwards, W. (1961). Probability learning in 1000 trials. *Journal of Experimental Psychology*, *62*(4), 385–394. https://doi.org/10.1037/h0041970

Even-Dar, E., Mannor, S., & Mansour, Y. (2006). Action Elimination and Stopping Conditions for the Multi-Armed Bandit and Reinforcement Learning Problems. *Journal of Machine Learning Research*, *7*(39), 1079–1105. Retrieved February 17, 2022, from http://jmlr.org/papers/v7/evendar06a.html

Fechner, G. (1860). *Elemente der Psychophysik*. Breitkopf und Härtel. https://books.google.de/books?id=6rINAAAAYAAJ

Gabillon, V., Ghavamzadeh, M., & Lazaric, A. (2012). Best Arm Identification: A Unified Approach to Fixed Budget and Fixed Confidence. *Advances in Neural Information Processing Systems*, *25*. Retrieved February 17, 2022, from https://proceedings.neurips.cc/paper/2012/hash/8b0d268963dd0cfb808aac48a549829f-Abstract.html

Gigerenzer, G., & Brighton, H. (2009). Homo Heuristicus: Why Biased Minds Make Better Inferences [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1756-8765.2008.01006.x]. *Topics in Cognitive Science*, *1*(1), 107–143. https://doi.org/10.1111/j.1756-8765.2008.01006.x

Gigerenzer, G., Hertwig, R., Van Den Broek, E., Fasolo, B., & Katsikopoulos, K. V. (2005). A 30% Chance of Rain Tomorrow: How Does the Public Understand Probabilistic Weather Forecasts? [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1539-6924.2005.00608.x]. *Risk Analysis*, *25*(3), 623–629. https://doi.org/10.1111/j.1539-6924.2005.00608.x

Gonzalez, C., & Mehlhorn, K. (2016). Framing From Experience: Cognitive Processes and Predictions of Risky Choice. *Cognitive Science*, *40*(5), 1163–1191. https://doi.org/10.1111/cogs.12268

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., ... Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments [Number: 1 Publisher: Nature Publishing Group]. *Scientific Data*, *3*(1), 160044. https://doi.org/10.1038/sdata.2016.44

Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description-experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making*, *21*(5), 493–518. https://doi.org/10.1002/bdm.598

Hertwig, R. (2015). Decisions from Experience. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell Handbook of Judgment and Decision Making* (pp. 239–267). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118468333.ch8

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from Experience and the Effect of Rare Events in Risky Choice. *Psychological Science*, *15*(8), 534–539. https://doi.org/10.1111/j.0956-7976.2004.00715.x

Hertwig, R., & Erev, I. (2009). The descriptionexperience gap in risky choice. *Trends in Cognitive Sciences*, *13*(12), 517–523. https://doi.org/10.1016/j.tics.2009.09.004

Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: Why small samples? *Cognition*, *115*(2), 225–237. https://doi.org/10.1016/j.cognition.2009.12.009

Hills, T. T., & Hertwig, R. (2010). Information Search in Decisions From Experience: Do Our Patterns of Sampling Foreshadow Our Decisions? *Psychological Science*, *21*(12), 1787–1792. https://doi.org/10.1177/0956797610387443

Holdgraf, C., Appelhoff, S., Bickel, S., Bouchard, K., D'Ambrosio, S., David, O., Devinsky, O., Dichter, B., flinker adeen, a., Foster, B., Gorgolewski, K. J., Groen, I. I., Groppe, D., Gunduz, A., Hamilton, L., Honey, C. J., Jas, M., Knight, R., Lachaux, J.-P., ... Hermes, D. (2018). *BIDS-iEEG: An extension to the brain imaging data structure (BIDS) specification for human intracranial electrophysiology* (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/r7vc2

Juechems, K., Balaguer, J., Spitzer, B., & Summerfield, C. (2021). Optimal utility and probability functions for agents with finite computational precision [Publisher: National Academy of Sciences Section: Social Sciences]. *Proceedings of the National Academy of Sciences*, *118*(2). https://doi.org/10.1073/pnas.2002232118

Kang, Z., & Spitzer, B. (2021). Concurrent visual working memory bias in sequential integration of approximate number. *Scientific Reports*, *11*(1), 5348. https://doi.org/10.1038/s41598-021-84232-7

Lejarraga, T., Hertwig, R., & Gonzalez, C. (2012). How choice ecology influences search in decisions from experience [Place: Netherlands Publisher: Elsevier Science]. *Cognition*, *124*(3), 334–342. https://doi.org/10.1016/j.cognition.2012.06.002

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources [Publisher: Cambridge University Press]. *Behavioral and Brain Sciences*, *43*. https://doi.org/10.1017/S0140525X1900061X

Ludvig, E. A., Madan, C. R., Pisklak, J. M., & Spetch, M. L. (2014). Reward context determines risky choice in pigeons and humans [Publisher: Royal Society]. *Biology Letters*, *10*(8), 20140451. https://doi.org/10.1098/rsbl.2014.0451

Madan, C. R., Ludvig, E. A., & Spetch, M. L. (2019). Comparative inspiration: From puzzles with pigeons to novel discoveries with humans in risky choice. *Behavioural Processes*, *160*, 10–19. https://doi.org/10.1016/j.beproc.2018.12.009

Markant, D., Pleskac, Timothy J., Diederich, Adele, Pachur, Thorsten, & Hertwig, Ralph. (2015). Modeling choice and search in decisions from experience: A sequential sampling approach. *Proceedings of the 37th Annual Conference of the Cognitive Science Society*.

Niso, G., Gorgolewski, K. J., Bock, E., Brooks, T. L., Flandin, G., Gramfort, A., Henson, R. N., Jas, M., Litvak, V., T. Moreau, J., Oostenveld, R., Schoffelen, J.-M., Tadel, F., Wexler, J., & Baillet, S. (2018). MEG-BIDS, the brain imaging data structure extended to magnetoencephalography. *Scientific Data*, *5*, 180110. https://doi.org/10.1038/sdata.2018.110

Norgaard, M., Matheson, G. J., Hansen, H. D., Thomas, A., Searle, G., Rizzo, G., Veronese, M., Giacomel, A., Yaqub, M., Tonietto, M., Funck, T., Gillman, A., Boniface, H., Routier, A., Dalenberg, J. R., Betthauser, T., Feingold, F., Markiewicz, C. J., Gorgolewski, K. J., ... Ganz, M. (2022). PET-BIDS, an extension to the brain imaging data structure for positron emission tomography. *Scientific Data*, *9*(1), 65. https://doi.org/10.1038/s41597-022-01164-1

Ostwald, D., Starke, L., & Hertwig, R. (2015). A normative inference approach for optimal sample sizes in decisions from experience. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01342

Pernet, C. R., Appelhoff, S., Gorgolewski, K. J., Flandin, G., Phillips, C., Delorme, A., & Oostenveld, R. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography [Number: 1 Publisher: Nature Publishing Group]. *Scientific Data*, *6*(1), 103. https://doi.org/10.1038/s41597-019-0104-8

Rakow, T., Demes, K. A., & Newell, B. R. (2008). Biased samples not mode of presentation: Re-examining the apparent underweighting of rare events in experience-based choice. *Organizational Behavior and Human Decision Processes*, *106*(2), 168–179. https://doi.org/10.1016/j.obhdp.2008.02.001

Rakow, T., & Newell, B. R. (2010). Degrees of uncertainty: An overview and framework for future research on experience-based choice. *Journal of Behavioral Decision Making*, *23*(1), 1–14. https://doi.org/10.1002/bdm.681

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, *58*(5), 527–535. https://doi.org/bams/1183517370

Shadlen, M. N., & Shohamy, D. (2016). Decision Making and Sequential Sampling from Memory. *Neuron*, *90*(5), 927–939. https://doi.org/10.1016/j.neuron.2016.04.036

Shahrampour, S., Noshad, M., & Tarokh, V. (2017). On Sequential Elimination Algorithms for Best-Arm Identification in Multi-Armed Bandits [arXiv: 1609.02606 version: 2]. *IEEE Transactions on Signal Processing*, *65*(16), 4281–4292. https://doi.org/10.1109/TSP.2017.2706192

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, *50*(3), 665–690. https://doi.org/10.1016/S0304-3932(03)00029-1

Spaniol, J., & Wegier, P. (2012). Decisions from Experience: Adaptive Information Search and Choice in Younger and Older Adults. *Frontiers in Neuroscience*, *6*. https://doi.org/10.3389/fnins.2012.00036

Spitzer, B., Waschke, L., & Summerfield, C. (2017). Selective overweighting of larger magnitudes during noisy numerical comparison. *Nature Human Behaviour*, *1*(8), 1–8. https://doi.org/10.1038/s41562-017-0145

Summerfield, C., & Tsetsos, K. (2015). Do humans make good decisions? *Trends in cognitive sciences*, *19*(1), 27–34. https://doi.org/10.1016/j.tics.2014.11.005

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (Second edition). The MIT Press.

Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal [ISBN: 9781119569107 Publisher: National Academy of Sciences Section: Biological Sciences]. *Proceedings of the National Academy of Sciences*, *109*(24), 9659–9664. https://doi.org/10.1073/pnas.1119569109

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/10.1007/BF00122574

Urai, A. E., de Gee, J. W., Tsetsos, K., & Donner, T. H. (2019). Choice history biases subsequent evidence accumulation (T. Verstynen, B. G. Shinn-Cunningham, & T. Verstynen, Eds.) [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *8*, e46331. https://doi.org/10.7554/eLife.46331

Weber, E. U., Shafir, S., & Blais, A.-R. (2004). Predicting Risk Sensitivity in Humans and Lower Animals: Risk as Variance or Coefficient of Variation. *Psychological Review*, *111*(2), 430–445. https://doi.org/10.1037/0033-295X.111.2.430

Wulff, D., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*, *144*(2), 140–176. https://doi.org/10.1037/bul0000115

Wyart, V., de Gardelle, V., Scholl, J., & Summerfield, C. (2012). Rhythmic Fluctuations in Evidence Accumulation during Decision Making in the Human Brain. *Neuron*, *76*(4), 847–858. https://doi.org/10.1016/j.neuron.2012.09.015

Wyart, V., Myers, N. E., & Summerfield, C. (2015). Neural Mechanisms of Human Perceptual Choice Under Focused and Divided Attention [Publisher: Society for Neuroscience Section: Articles]. *Journal of Neuroscience*, *35*(8), 3485–3498. https://doi.org/10.1523/JNEUROSCI.3276-14.2015

2 | Control over sampling boosts numerical evidence processing in human decisions from experience

Stefan Appelhoff, Ralph Hertwig & Bernhard Spitzer

Appelhoff, **S.**, Hertwig, R., & Spitzer, B. (2022). Control over sampling boosts numerical evidence processing in human decisions from experience. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhac062

Abstract

When acquiring information about choice alternatives, decision makers may have varying levels of control over which and how much information they sample before making a choice. How does control over information acquisition affect the quality of sample-based decisions? Here, combining variants of a numerical sampling task with neural recordings, we show that control over when to stop sampling can enhance (i) behavioral choice accuracy, (ii) the build-up of parietal decision signals, and (iii) the encoding of numerical sample information in multivariate electroencephalogram patterns. None of these effects were observed when participants could only control which alternatives to sample, but not when to stop sampling. Furthermore, levels of control had no effect on early sensory signals or on the extent to which sample information leaked from memory. The results indicate that freedom to stop sampling can amplify decisional evidence processing from the outset of information acquisition and lead to more accurate choices.

2.1 Introduction

Humans routinely acquire information about choice alternatives before deciding between them. In many situations, decision makers can control which and how much information they sample. For example, when deciding which of 2 products to buy, a customer may deliberately study reviews and testimonials before making a final choice. In other situations, the availability and amount of relevant information is determined by external factors. For instance, when selecting job applicants in an organization that uses standardized interviews, an employer must decide based on the applicants' answers to the same set of predefined questions. More generally, decision scenarios can differ in the extent to which an agent has control over sampling, in terms of which and how much information is sampled before a choice is made.

One experimental setup suitable for studying how control over sampling may affect decision-making is a numerical sampling paradigm (Hertwig et al., 2004; Hertwig & Erev, 2009) in which participants can view sequential samples of possible choice outcomes before deciding for one or the other option. The paradigm has been used extensively in behavioral studies of risky choice to examine how decision makers choose between options about which they learned from experience (i.e., through sampling the payoff distribution; "experience-based" decisions) as opposed to from formal description (where participants would be explicitly informed that there is, e.g., "25% chance to obtain €10, otherwise €0"; Hertwig, 2015; Wulff et al., 2018) Across these studies, researchers have also varied the extent to which participants were able to control the sampling process themselves. While the standard paradigm allows participants to decide freely which alternatives to sample and how often (Hertwig & Erev, 2009), some studies have prespecified the total number of samples to be taken (Fleischhut et al., 2014; Gonzalez & Mehlhorn, 2016; Hau et al., 2008; Ungemach et al., 2009) or included matched ("yoked") conditions in which participants had no control at all over the sampling sequence (Rakow et al., 2008). However, the latter variants of the sampling paradigm have been devised primarily to reduce confounds in comparison with decisions from description (Rakow & Newell, 2010). Therefore, it still remains unclear how control over sampling may alter experience-based decision-making itself.

Several lines of evidence suggest that a sense of control can be beneficial in cognitive tasks (Gureckis & Markant, 2012; Murayama et al., 2016). Agency in information acquisition has, for instance, been found to improve subsequent memory performance (Voss et al., 2011), even when exposure to the information was held constant (Murty et al., 2015). Another line of work has shown better performance in tasks self-selected by the participant than when the same tasks were selected by an experimenter (Murayama et al., 2015). More generally, various studies have identified performance benefits associated with volitional control per se and indicated that such effects could be mediated by motivational factors (Patall, 2012; Patall et al., 2008). However, the effects of control cannot easily be generalized across domains. In some contexts, control does not seem to impact task performance (Flowerday & Schraw, 2003; Flowerday et al., 2004) or can be detrimental – for instance, when control is perceived as irrelevant or as too complex (Katz & Assor, 2007; but see Murayama et al., 2015).

In the domain of decisions from experience using the sampling paradigm (Hertwig et al., 2004; Hertwig & Erev, 2009), understanding of the role of agency in the sampling process is rather incomplete. In a recent meta-analysis, Wulff et al. (2018) suggested that control over sampling appears to alter the temporal weighting of numerical samples in subsequent choice. Specifically, when participants were given full control over sampling, their choices indicated stronger "recency"

effects (i.e., a tendency to overweight the later samples in a sequence, which is routinely observed in sequential tasks with discrete samples, e.g., Anderson, 1964; Cheadle et al., 2014; Kang & Spitzer, 2021; Spitzer et al., 2017; Tsetsos et al., 2012; D. J. Weiss & Anderson, 1969; Wyart et al., 2015). However, the meta-analysis by Wulff et al. (2018) was limited to comparisons across studies and did not address the general performance benefits (or drawbacks) that may be associated with control over sampling or the neurocognitive processes that might underlie them.

Here, we used specially designed variants of a numerical sampling paradigm combined with Electroencephalography (EEG) recordings to study how control over sampling affects experience-based decision-making. We systematically varied whether participants (i) were free to decide how much information to sample and from which option (full control), or (ii) could decide only from which option to sample but with a prespecified total number of samples (partial control), or (iii) had no control over sampling at all (no control). Importantly, our design controlled for differences in stimulus presentation by matching the sample sequences in the no-control conditions with those in the self-controlled tasks (full or partial control). We found that full, but not partial, control over sampling had a beneficial effect on choice accuracy and that this benefit was associated with a stronger encoding of numerical sample information from the outset of information acquisition.

2.2 Materials and Methods

2.2.1 Participants

Forty healthy volunteers took part in the experiment (20 female, 20 male; mean age 26.3 ± 3.7 years; all right-handed). All participants provided written informed consent and received a flat fee of €10 and €10 per hour as compensation, as well as a performance-dependent bonus (€9.35 ± €0.48 on average). The study was approved by the ethics committee of the Max Planck Institute for Human Development.

2.2.2 Experimental design

The tasks were variants of the classic sampling paradigm described in Hertwig and Erev (2009). On each trial, in all experimental conditions, participants were asked to decide between 2 choice options (left or right), each of which could return 1 of 2 different reward values (displayed as an Arabic digit between 1 and 9; Figure 2.1*a*, *green digit*). Prior to making a final decision for one of the options, participants viewed samples from each option. That is, they could preview potential choice outcomes (Figure 2.1*a*, *white digits*). Each option returned one outcome (e.g., "1") with probability p, and another outcome (e.g., "9") with probability 1-p. The outcome probability p of the options ranged from 0.1 to 0.9 (in steps of 0.1) and remained constant in the course of a trial. We constrained the outcome values and probabilities on each trial such that (i) none of the 4 possible outcome values (2 for each option) were identical and (ii) the difference in expected value between the 2 options was always 0.9 (based on piloting results). Under these constraints, the choice problems presented on each trial were selected pseudorandomly, with the additional restriction that each sample value (1, 2, ..., 9) occurred with approximately equal probability across the experiment. Participants were instructed to learn from the observed samples and to finally select the option that they expected to return the higher reward (i.e., the larger numerical value). Participants were told that the reward returned by their final choice would influence their bonus payout at the end of the experiment.

Half of the participants were assigned to the "full control" condition, where they were free to

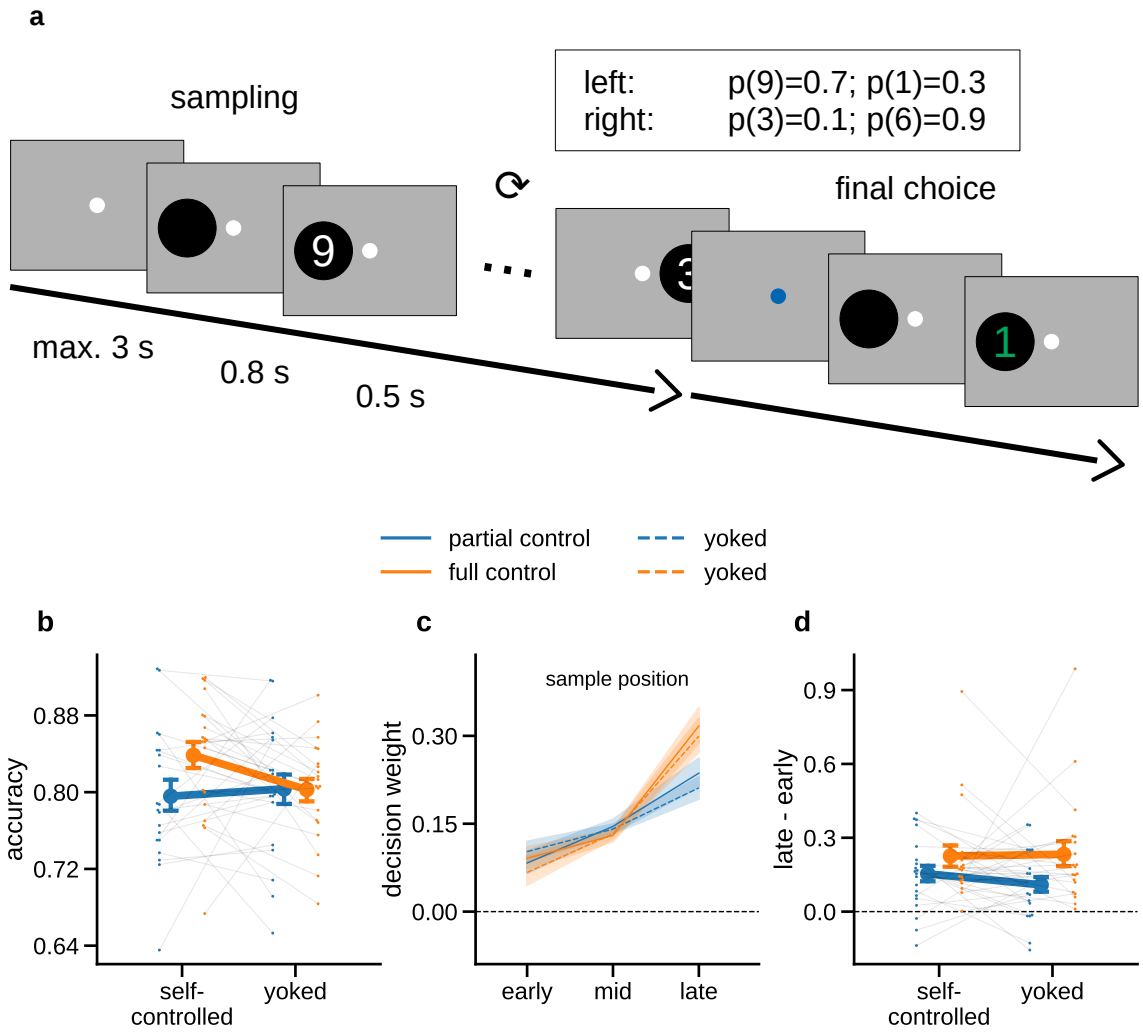

Figure 2.1. Experimental task and behavioral results. a) Schematic illustration of an example trial. Participants were asked to decide which of 2 choice options (left or right) would yield a larger numerical outcome. Before committing to a choice, participants could draw up to 19 samples (full control group) or were required to draw a fixed number of 12 samples (partial control group). Samples are shown as white digits; the final choice outcome is shown in green. The inset table shows the outcome values and probabilities for the 2 choice options in the example trial. In yoked baseline conditions, participants judged replays of previously recorded sampling streams. **b)** Mean accuracy (proportion of times the option with the higher mean of samples was chosen) in each condition. **c)** Decision weights (see Materials and Methods) of samples occurring early, mid, or late in the sampling sequence, for each sampling condition. **d)** Difference in decision weight between late and early samples. Higher values indicate that late samples had a stronger relative influence on choice than early samples ("recency" effect). Error indicators in all panels show SE.

sample from the left or right option as often as they wished before making a final choice. The only restriction on sampling in the full control condition was that a sample had to be taken within 3 s (otherwise the trial was restarted) and that the total number of samples could not exceed 19. The other half of participants were assigned to the "partial control" condition, which was identical to the full control condition except that a fixed number of 12 samples had to be drawn on every trial. The number of samples was based on pilot data where free-sampling participants took approximately 12 samples on average. In other words, participants in the partial control condition were also free to sample from the left or right option but had no control over when to stop (or continue) sampling: They were always prompted to make a final choice after the 12th sample.

In both sampling conditions, the beginning of a new trial was signaled by a green fixation stimulus

(a combination of bull's-eye and cross hair; Thaler et al., 2013) that turned white after 1 s. Upon pressing the left or right button on a USB response pad (using the left or right hand, respectively), participants were shown a black circular disk (diameter 5° visual angle) 4.5° to the left (choice option 1) or right (choice option 2) of fixation after 0.2-0.4 s (randomly varied). After another delay of 0.8 s, the number sample was presented in white (font Liberation Sans, height 4°) in the disk area for 0.5 s (see Figure 2.1*a* for a schematic illustration). After this, the disk disappeared and participants were given 3 s to draw the next sample. The black disk served as a spatial cue to minimize differences in surprise about the sample location (left/right) in yoked conditions without sampling control (see below). The sampling procedure was repeated depending on condition (partial control: 12 samples; full control: up to 19 samples), and the resulting sample sequences (including their precise timing) were recorded (see yoked conditions below). In the full control condition, a third button on the response pad (above the "right" button) was available to stop the sampling sequence. In all conditions, after the sampling was finished, the fixation stimulus changed color to blue for 1 s and participants were asked to make a final choice between the left and right options. The button and display procedure for the final choice was identical to that for drawing samples, except that the final choice outcome was displayed in green to indicate the eventually obtained reward. The rewards (i.e., the payouts from the final choices) were converted to Euros with a factor of 0.005 and added as a bonus to participant's reimbursement after the experiment (see Participants above).

Within both groups (full and partial control), each participant additionally performed the task in a "yoked" condition, where they had no control over sampling. Here, participants made decisions based on replays of previously recorded sampling streams (without any control over which and how many samples were shown or their timing). Accordingly, we refer to the yoked conditions as the no-control baseline conditions. In each group, half of the participants first performed the self-controlled sampling task (full or partial) and subsequently performed the no-control task with a replay of their own sampling sequences. In informal debriefing after the experiment, none of these participants reported to have noticed that they viewed exact replays of their own sampling sequences. The other half of the participants in each group performed the no-control task first (yoked to the sampling sequences of another participant in the same group) and the respective self-controlled task second. Control analysis showed no differences in choice accuracy between participants who performed the baseline task first (yoked to another participant's sequences) or second (yoked to their own sequences) (all $p > 0.05$). Furthermore, in the subset of participants who were yoked to another participant, the difficulty of active versus yoked sampling sequences did not differ (all $p > 0.05$). Each participant performed 100 trials (5 blocks of 20 trials with short breaks between blocks) in the self-controlled and yoked task variant, respectively.

Participants in the full control group drew on average 8.6 samples (SD=4.2, median=8), compared with the 12 samples that had to be drawn in the partial control group. Due to the principled impossibility of matching full and partial control trials (e.g., with respect to the precise length and timing of the sampling sequences on individual trials), all our analyses focus on comparisons of differences to the yoked baseline condition within each group. This analysis strategy rules out stimulus confounds that may arise, for instance due to "amplification effects" under full control, where the decision to stop sampling may be more likely when the momentary difference between the accumulated option values happens to be large (Hertwig & Pleskac, 2010).

The experiment was programmed in Python using the PsychoPy package (Peirce et al., 2019) and run on a Windows 10 PC. The experiment code is available on Zenodo (https://doi.org/10.

5281/zenodo.3354368). Behavioral responses were recorded using a USB response pad (The Black Box ToolKit Ltd, United Kingdom). Throughout the experiment, eye movements were recorded using a Tobii 4C Eye-Tracker (Tobii Technology, Sweden; sampling rate 90 Hz). To reduce eye movements, participants' gaze position was analyzed online while the experiment was run in all sampling conditions. The program displayed a warning message and restarted the trial whenever the gaze left an elliptical area centered on the central fixation stimulus (width 5° visual angle, height 2.85° visual angle) more than 4 times during a trial. Saccades towards the outcome samples were robustly detected with these settings. On average, 3% of trials per participant were restarted due to a lack of fixation or failure to draw a sample within 3 s (see above). Offline analyses confirmed that participants generally held fixation in the remaining trials.

2.2.3 Supplementary tasks

After the main experiment, participants performed an additional short task on the same choice problems, where the options were not explored through sampling but described formally on screen (e.g., "8 with 60% or 4 with 40%?"). Due to a coding error, much of the data (84%) from this task was incorrectly recorded and the results are thus not reported here. Participants further completed a brief numeracy questionnaire (Berlin Numeracy Test (BNT); Cokely et al., 2012). Exploratory analysis showed no significant correlations of the effects reported in our main analysis with BNT scores.

2.2.4 EEG recording

The experiment was performed in an electrically shielded and soundproof cabin. Scalp EEG was recorded with 64 active electrodes (actiCap, Brain Products GmbH Munich, Germany) positioned according to the international 10% system. Electrode FCz was used as the recording reference. We additionally recorded the horizontal and vertical electrooculogram (EOG) and electrocardiogram (ECG) using passive electrode pairs with bipolar referencing. All electrodes were prepared to have an impedance of less than 10 kΩ. The data were recorded using a BrainAmp DC amplifier (Brain Products GmbH Munich, Germany) at a sampling rate of 1,000 Hz, with an RC high-pass filter with a half-amplitude cutoff at 0.016 Hz (roll-off: 6 dB/octave) and low-pass filtered with an anti-aliasing filter of half-amplitude cutoff 450 Hz (roll-off: 24 dB/octave). The dataset is organized in Brain Imaging Data Structure format (BIDS; Gorgolewski et al., 2016) according to the EEG extension (Pernet et al., 2019) and is available on GIN (https://doi.org/10.12751/g-node.dtyh14).

2.2.5 Behavioral data analysis

We quantified choice accuracy in each condition as the proportion of trials on which participants chose the option in which the observed samples were on average larger. A choice was thus defined as correct when the experienced samples of the chosen option had the higher arithmetic mean. This choice corresponds to that of a noiseless ideal observer in the task, given the presented samples. Differences in accuracy between sampling conditions were analyzed using a mixed 2x2 analysis of variance (ANOVA, self-controlled/yoked; full/partial), followed up with Bonferroni-corrected pairwise t-tests. All statistical tests reported (including in the EEG analyses, see below) are 2-tailed.

Based on previous work, we expected participants to show a recency effect, that is, a relative overweighting of later samples. To quantify recency effects in the behavioral data, we used a reverse correlation approach (Neri et al., 1999; Spitzer et al., 2016) based on logistic regression.

We first divided the samples in a trial into early, mid, and late samples. The first and last 2 samples in a trial were defined as early and late samples, respectively, and the remaining samples as "mid" samples. Trials with fewer than 5 samples overall were discarded in this analysis (between 1% and 41.5% of trials per participant, mean=13.3%). For each participant, task condition, and time window, we regressed the participant's final choices (left: 0, right: 1) onto the numerical sample values (numbers 1, 2, ..., 9 rescaled to -4, -3, ..., 4), where the values for the left option were sign-flipped to reflect their opposite impact on the probability of choosing the right option (Spitzer et al., 2017). The regression coefficients resulting from this analysis provide a measure of "decision weight", that is, of the influence that number samples (early, mid, or late) had on choice. We quantified recency as the difference in weight between late and early samples Figure 2.1*d*. Differences in recency between conditions were assessed with a 2x2 ANOVA specified analogously as above.

2.2.6 EEG preprocessing

The EEG recordings were visually inspected for noisy segments and bad channels. Ocular and cardiac artifacts were corrected using independent component analysis (ICA). To this end, we high-pass filtered a copy of the raw data at 1 Hz and downsampled it to 250 Hz. We then ran an extended infomax ICA on all EEG channels and time points that were not marked as bad in the prior inspection. Using the EOG and ECG recordings, we identified stereotypical eyeblink, eye movement, and heartbeat artifact components through correlation with the independent component time courses. We visually inspected and rejected the artifact components before applying the ICA solution to the original raw data (Winkler et al., 2015). We then filtered the ICA-cleaned data between 0.1 and 40 Hz, interpolated bad channels, and re-referenced each channel to the average of all channels. Next, the data were epoched from -0.2 to 0.8 s relative to each individual number sample onset. Remaining bad epochs were rejected using a thresholding approach from the FASTER pipeline (Step 2; Nolan et al., 2010). On average, n=1925 clean epochs (93.85%) per participant were retained for analysis. The epochs were downsampled to 250 Hz and baseline corrected relative to the period from -0.2 to 0 s before stimulus onset. All EEG analyses were performed in Python using MNE-Python (Gramfort et al., 2013), MNE-BIDS (Appelhoff et al., 2019), and custom code. All analysis code is available on Zenodo (https://doi.org/10.5281/zenodo.5929222).

2.2.7 Event-related potential analysis

EEG analyses are reported for the epochs around the onset of the individual number samples. We first examined lateralized visual Event Related Potential (ERP) components to test whether early visual processing differed between the sampling conditions. To this end, we subtracted the ERP for stimuli presented on the right from the ERP for stimuli presented on the left and then subtracted the mean signal of right-hemispheric (O2, PO4, PO8, PO10) occipitoparietal channels of interest (based on previous literature; Eimer, 1998) from the corresponding left-hemispheric (O1, PO3, PO7, PO9) channels. Mean amplitudes of the lateralized evoked potential were extracted from prototypical time windows (P1 ERP component: 80-130 ms, N1 ERP component: 140-200 ms) for each sampling condition and analyzed in a mixed 2x2 ANOVA (self-controlled/yoked; full/partial).

We further examined centro-parietal evoked responses (CPP/P3, averaged over the early, mid, and late samples in each trial) as a potential correlate of decisional evidence accumulation (O'Connell et al., 2012; Pisauro et al., 2017; Twomey et al., 2015). To this end, we averaged the signal over centro-parietal channels (Cz, C1, C2, CPz, CP1, CP2, CP3, CP4, Pz, P1, P2) and focused on a

time window from 300 to 600 ms, based on previous analyses of CPP/P3 responses during visual stimulus sequences (Polich, 2007; Spitzer et al., 2017; Wyart et al., 2015).

2.2.8 Representational similarity analysis

To examine the encoding of numerical sample value in multivariate ERP patterns, we used an approach based on representational similarity analysis (RSA, Kriegeskorte & Kievit, 2013). For RSA, the ERPs were additionally smoothed (Grootswagers et al., 2016) with a Gaussian kernel (35 ms half duration at half maximum). We used a conventional ERP-RSA approach (e.g., Luyckx et al., 2019; Spitzer et al., 2017), where the representational geometry of a stimulus space (here, sample values 1-9) is characterized by the multivariate (dis-)similarity between the ERP topographies (comprising all 64 channels) associated with each sample value. Representational dissimilarity was computed at each time point of the ERP, between each pair of stimuli (using Euclidean distance as dissimilarity measure), yielding a 9x9 representational dissimilarity matrix (RDM; see Figure 2.3*a*, *lower*) at each time point. We refer to the RDMs computed from the ERP data as "ERP-RDMs".

To the extent that multivariate ERP patterns encode numerical sample information, they should show a "numerical distance" effect (e.g., Luyckx et al., 2019; Spitzer et al., 2017; Teichmann et al., 2018). That is, the representational dissimilarity between, for example, numbers "2" and "3" should be smaller than that between "2" and "4", which, in turn, should be smaller than that between "1" and "4", and so forth, for any pairing of numbers. To assess numerical distance effects in our ERP-RDMs, we created a theoretical model RDM (Figure 2.3*a*, *upper*) where each cell reflects the actual numerical difference between sample values (i.e., the numerical distance between "3" and "7" is 4, and that between "4" and "6" is 2). We then quantified the match between the model RDM and the ERP-RDM at each time point by computing the correlation (Pearson's r) between the two, with stronger correlation indicating stronger encoding of numerical magnitude in multivariate ERP patterns (see also Spitzer et al., 2017; Teichmann et al., 2018). Correlations between model- and ERP-RDMs were restricted to the lower triangle (excluding the diagonal) to omit redundant matrix entries.

In addition to numerical distance, we examined the extent to which ERP patterns encoded the "extremity" of a sample value (i.e., its absolute difference from the midpoint of the sample range, 5; Figure 2.3*d*). To avoid confounds by potential deviations from a uniform distribution of sample values across the experiment, we additionally orthogonalized each model RDM to an RDM reflecting the relative frequency of numerical sample occurrences (see also Spitzer et al., 2017). However, qualitatively similar results were obtained when this orthogonalization step was omitted.

For statistical analysis, we used t-tests against zero with cluster-based permutation testing (Maris & Oostenveld, 2007) to control for multiple comparisons over time points (10,000 iterations, cluster-defining threshold p=0.05). We then recomputed the ERP-RSA separately for each sampling condition to test for differences in number encoding. Differences between conditions were examined using mixed 2x2 ANOVAs (self-controlled/yoked; full/partial), again using cluster-based permutation testing to control for multiple comparisons over time points. Analogous RSA analyses were performed separately on the first and second half of samples from each trial (Figure 2.3*c*).

2.2.9 Analysis of neurometric distortions

Our basic RSA of numerical distance and extremity assumes a linear representation of numerical magnitude, where the representational distance between, for example, "3" and "5" is the same as that between e.g., "7" and "9". However, based on previous work (e.g., Luyckx et al., 2019; Nieder,

2016; Spitzer et al., 2017), the neural representation of numbers might be nonlinearly distorted. That is, the neural number representation might be compressed (such that the representational distance between e.g., "8" and "9" is smaller than that between "5" and "6") or anti-compressed (such that the distance between "8" and "9" is larger). To examine such potential distortions (see also Spitzer et al., 2017), we transformed the numerical sample values using a parameterized power function $v = sign(x + b) \times |x + b|^k$, where x are the numerical sample values (1-9 normalized to the range [-1, 1]), exponent k determines the shape of the transformation ($k < 1$ compression, $k = 1$ linear, $k > 1$ anti-compression; see inset plot in Figure 2.4*b* for illustration of the resulting distortions), and b reflects an overall bias towards smaller ($b < 0$) or larger numbers ($b > 0$). We then created model RDMs (analogously as above) from the thus transformed values (v), for different values of k (varied between 0.5 and 10) and b (varied between -0.75 and 0.75; where $k = 1$ and $b = 0$ corresponds to linear/unbiased transformation). For each parameter combination, we correlated the resulting model RDM with the ERP-RDM (analogously as above). In each participant, the parameter combination for which the model RDM correlated most strongly with the ERP-RDM was used as the estimate of the participant's neurometric distortion. Statistical analysis of the neurometric parameters proceeded with conventional statistical tests on the group level.

2.3 Results

Participants (n=40) observed sequential samples (Arabic digits 1-9) of the potential rewards of choice options (left/right) before deciding on one of them Figure 2.1*a*. In different conditions, participants (i) could determine from which option(s) to sample and when to stop sampling ("full control", 1-19 samples/trial, n=20 participants) or (ii) could determine only from which option to sample for a fixed number of samples ("partial control", 12 samples/trial, n=20 participants). Each participant additionally performed the task in a "yoked" condition with matched sample sequences (see Materials and Methods) that they could not control. Our behavioral and EEG analyses focus on the effects of control (full or partial) relative to the respective matched (yoked) no-control conditions.

2.3.1 Behavior

Mean choice accuracy (i.e., the percentage of trials on which participants chose the option in which the average of the sampled values was larger, see Materials and Methods) was 83.8% under full control (SE=1.4%, yoked baseline: 80.3%, SE=1.2%) and 79.6% under partial control (SE=1.6%, yoked baseline: 80.3%, SE=1.6%). A mixed 2x2 ANOVA with the factors control over sampling (self-controlled or yoked; within participants) and control type (full or partial; between participants) showed no main effects [self-controlled/yoked: $F(1,38)=2.143$, $p=0.151$, $\eta_p^2=0.053$; full/partial: $F(1,38)=1.321$, $p=0.258$, $\eta_p^2=0.034$], but a significant interaction of the 2 factors [$F(1,38)=5.108$, $p=0.03$, $\eta_p^2=0.118$]. Post hoc tests showed significantly higher accuracy under full control than in the yoked baseline [$t(19)=2.644$, $p=0.032$, $d=0.605$, Bonferroni corrected], but no such effect under partial control [$t(19)=-0.561$, $p>0.9$, $d=-0.108$]. Thus, relative to matched baseline conditions, we found an accuracy benefit of control over sampling under full control but not under partial control.

We next examined whether and how the temporal weighting of sample information differed between conditions. To this end, we examined the samples' decision weights (see Behavioral data analysis) separately for early, mid-, and late portions of the sampling sequence (Figure 2.1*c*). As expected

based on previous work (Anderson, 1964; Cheadle et al., 2014; Kang & Spitzer, 2021; Spitzer et al., 2017; Tsetsos et al., 2012; D. J. Weiss & Anderson, 1969; Wyart et al., 2015), we found a pronounced recency pattern, with decision weight generally increasing over the course of the trial. In other words, later samples had a higher impact on the final choice than earlier samples. For comparison between sampling conditions, we quantified recency as the difference in decision weight between late and early samples (Figure 2.1*d*). A mixed 2x2 ANOVA, specified analogously as for accuracy above, showed no significant main effects [self-controlled/yoked: $F(1,38)=0.8$, $p=0.377$, $\eta_p^2=0.021$; full/partial: $F(1,38)=3.363$, $p=0.075$, $\eta_p^2=0.081$] and no interaction between the 2 factors [$F(1,38)=1.483$, $p=0.231$, $\eta_p^2=0.038$]. Thus, we found no impact of control over sampling on recency. To summarize the behavioral results, full control over sampling was characterized by increased choice accuracy but was not distinguished in the extent to which sample information "leaked" (Usher & McClelland, 2001), or was forgotten, in the course of a trial.

2.3.2 Visual evoked responses

Turning to the EEG data, we first examined visual evoked responses to test whether the sampling conditions differed in terms of early sensory processing of the sample stimuli (e.g., due to potential differences in stimulus-directed visual attention; Luck et al., 2000). Figure 2.2*a* shows the occipitoparietal ERP difference between stimuli occurring in the right and left visual fields, subtracted between contralateral channels (see Event-related potential analysis). Statistical analysis showed no differences between sampling conditions in the time window of either the P1 (80-130 ms) or the N1 component (140-200 ms) of the visual ERP [all $F(1,38)<1.71$, all $p>0.20$, all $\eta_p^2<0.044$; mixed 2x2 ANOVAs specified as in the behavioral analysis above]. We thus found no evidence for differences in early visual processing between the sampling conditions.

2.3.3 Centro-parietal positivity/P3

We next examined centro-parietal positivity (CPP) responses over centro-parietal channels between 300 and 600 ms after stimulus onset. The amplitude of the CPP response to a sample generally increased in the course of the trial (Figure 2.2*b-d*), which is in line with previous studies implicating the CPP in decision formation (O'Connell et al., 2012; Twomey et al., 2015). Figure 2.2*c* and *d* illustrates the monotonic ramping up of CPP across samples occurring early, mid, and late in the trial (see Materials and Methods) under partial and full control. Descriptively, the build-up of CPP appeared stronger under full control. For statistical analysis, we examined the increase in CPP amplitude from early to late samples in the individual sampling conditions (Figure 2.2*b*). A significant increase in amplitude was evident in each condition (including yoked; Figure 2.2*b*, all $p<0.02$, t-tests against zero, uncorrected). A mixed 2x2 ANOVA comparing the amplitude difference between conditions showed no significant main effects [self-controlled/yoked: $F(1,38)=1.579$, $p=0.217$ $\eta_p^2=0.04$; full/partial: $F(1,38)=1.534$, $p=0.223$, $\eta_p^2=0.039$], but a significant interaction [$F(1,38)=11.408$, $p=0.002$, $\eta_p^2=0.231$]. Post hoc t-tests showed that the CPP increased more steeply in the full control condition than in the yoked baseline [$t(19)=2.772$, $p=0.024$, $d=0.687$, paired t-test, corrected], whereas no such effect was evident under partial control [$t(19)=-1.932$, $p=0.137$, $d=-0.355$]. Thus, the increased choice accuracy under full control was accompanied by a steeper increase of centro-parietal decision signals within trials (O'Connell et al. 2012; Twomey et al. 2015; Wyart et al. 2015; Spitzer et al. 2016). (O'Connell et al., 2012; Spitzer et al., 2016; Twomey et al., 2015; Wyart et al., 2015). Importantly, these effects were observed in comparison against matched (yoked) baseline, ruling out that they were attributable to any specific characteristics of the self-sampled stimulus sequences.

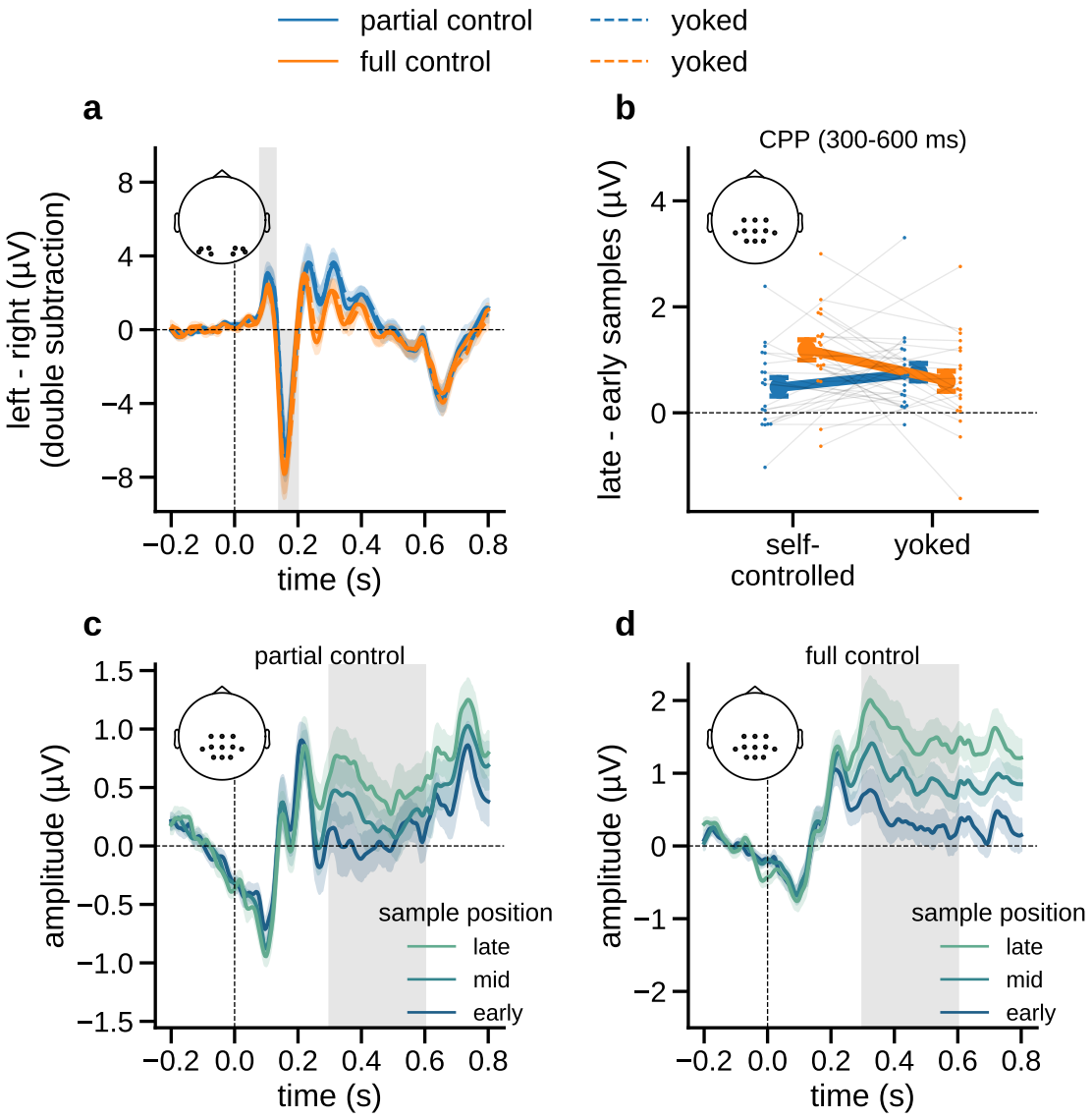

Figure 2.2. Univariate EEG results with ERPs time-locked to number sample onset. a) Early visual ERPs (left - right stimuli, right channels subtracted from left channels) in each sampling condition. Gray shadings indicate time windows of the P1 and N1 components, respectively (80-130 ms and 140-200 ms). **b)** The difference in centro-parietal (CPP) amplitudes between samples occurring late versus early in the trial (see panels *c* and *d*), plotted separately for each sampling condition (including yoked). **c)** The "ramping up" of CPP amplitudes (0.3-0.6 s) over early, mid, and late samples in the partial control condition. Gray shadings indicate the time window from which average amplitudes were extracted in panel *b*. **d)** Same as *c*, for the full control condition. Error indicators in all panels show SE.

2.3.4 Representational similarity analysis

Our results so far show that decisions made with full control over sampling were more accurate and accompanied by a stronger build-up of parietal choice signals (Figure 2.2*c*-*d*), whereas there were no differences in early visual processing (Figure 2.2*a*) or in the extent to which sample information "leaked" over time (i.e., no difference in recency effects; Figure 2.1*b*-*c*). One possibility is that a benefit of full control may have arisen at the stage of numerical processing, in encoding a sample's abstract value (i.e., its numerical magnitude, which is to be integrated into the subjectively perceived value of the choice option). We used an RSA-based approach (see Materials and Methods) to examine the neural encoding of the samples' numerical magnitude, building on previous findings

of numerical distance effects in multivariate ERP patterns (Luyckx et al., 2019; Sheahan et al., 2021; Spitzer et al., 2017; Teichmann et al., 2018). Specifically, we correlated the multivariate similarity structure of samples (1-9) in our ERP data with theoretical models reflecting (i) numerical distance and (ii) extremity of the sample values (see Materials and Methods).

Numerical distance

We found robust encoding of numerical magnitude in terms of a numerical distance effect in multivariate ERP signals between approximately 160 and 800 ms after sample onset (Figure 2.3*b*, $p_{cluster}<0.001$, t-test against zero), which replicates and extends previous findings in tasks without sampling control (Luyckx et al., 2019; Sheahan et al., 2021; Spitzer et al., 2017; Teichmann et al., 2018). To test whether the strength of this effect differed between levels of sampling control, we examined its time course in the various conditions (full, partial, yoked baselines) using mixed 2x2 ANOVAs (specified analogously as above). The analysis showed no main effects (all $p_{cluster}>0.05$) but a significant interaction cluster between 320 and 580 ms ($p_{cluster}=0.009$). We next compared the average numerical distance effects in the time window of this cluster. We found the effect to be significantly larger (relative to yoked baseline) under full control [t(19)=3.65, p=0.003, d=1.05, corrected] but not under partial control [t(19)=-1.065, p=0.6, d=-0.340, corrected]. In other words, the encoding of numerical magnitude in sample-level neural signals was enhanced under full control, mirroring the pattern of findings for CPP build-up (Figure 2.2*b*) and choice accuracy (Figure 2.1*b*).

We next asked whether the enhanced number encoding under full control was driven only by late samples occurring near the time of the decision to stop sampling. To this end, we repeated the RSA analysis separately for the first (Figure 2.3*c*, *left*) and second (Figure 2.3*c*, *right*) half of the samples in a trial. Importantly, a significant enhancement under full control relative to yoked baseline was already evident in the first half of samples [t(19)=2.279, p=0.034, d=0.707], that is, long before participants stopped sampling. The effect in the second half of samples was similar [t(19)=2.237, p=0.037, d=0.673; partial control: both p>0.24]. In sum, we found no indication that enhanced number encoding under full control occurred only near the time of deliberate (vs. forced) stopping. Rather, the effect appeared to emerge early in the sampling sequence. We note again that we only interpret effects in relation to the respective matched (yoked) control conditions, as other comparisons may suffer from nontrivial stimulus confounds (see Experimental design).

Extremity

Inspection of the empirically observed ERP-RDM (Figure 2.3*a*, *lower*) suggests that besides numerical distance, the multivariate ERP patterns also encoded the extremity of the sample values (i.e., their absolute distance from the midpoint of the sample range, see also Spitzer et al. 2017; Luyckx et al. 2019). Using a model RDM of numerical extremity (Figure 2.3*d*; note that the model is orthogonal to the numerical distance RDM in Figure 2.3*a*, *upper*), we found a significant effect between approximately 260 and 800 ms (t-test against zero, $p_{cluster}<0.001$) in the ERP data collapsed across conditions. However, testing for differences between sampling conditions yielded no significant results (all $p_{cluster}>0.05$). Together, while both numerical distance and numerical extremity were reflected in the multivariate ERP data, only numerical distance mirrored the enhancement under full control that was observed in CPP build-up and in behavior.

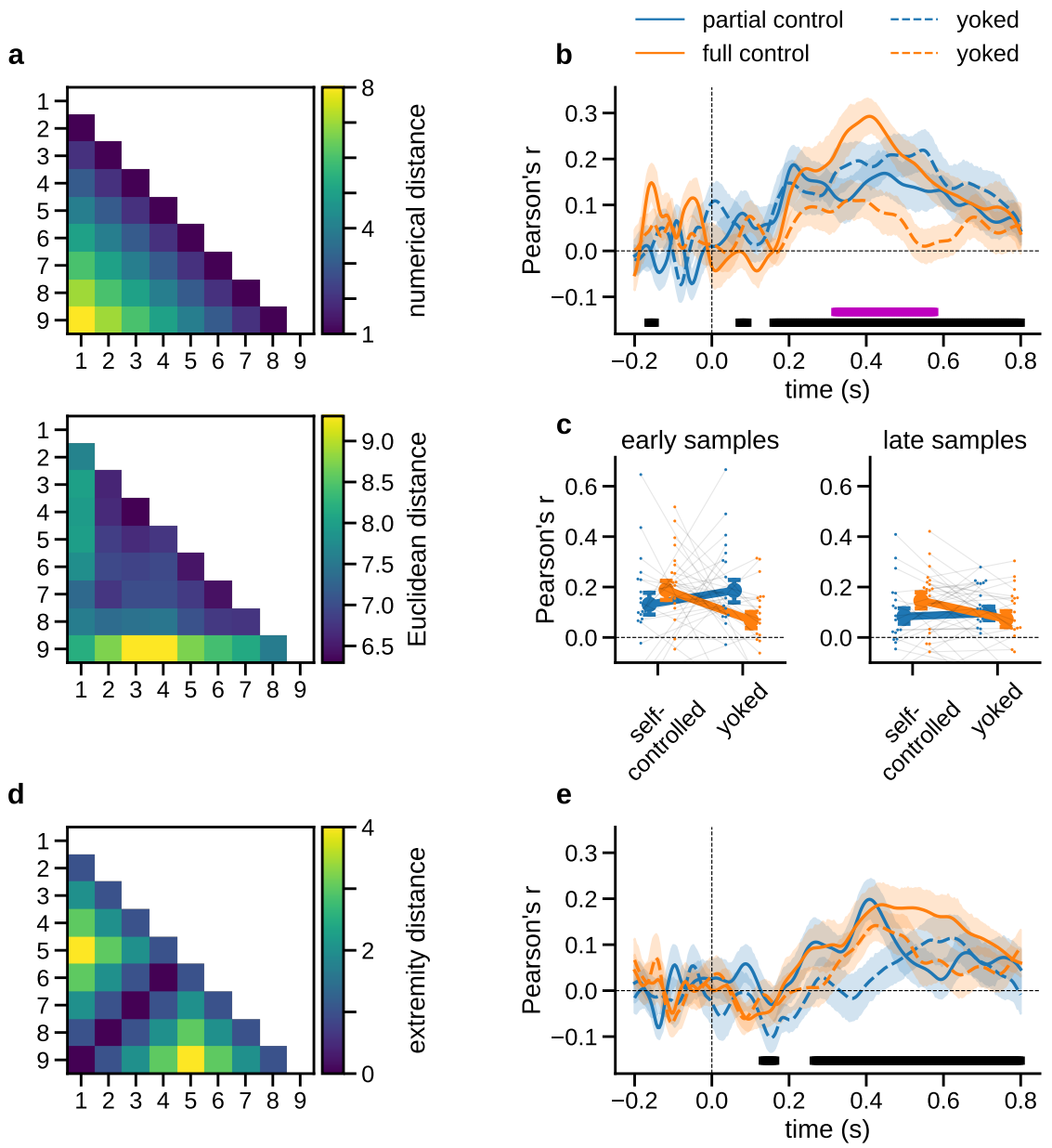

Figure 2.3. RSA results. a) *Upper:* Model RDM reflecting the pairwise numerical distance between sample values. *Lower:* Grand mean ERP-RDM averaged across participants and sampling conditions in a representative time window between 300 and 600 ms after sample onset. **b)** Time course of numerical distance effects in multivariate ERP patterns, plotted separately for each sampling condition. Black bar indicates time windows of significant numerical distance encoding (collapsed across sampling conditions). Purple bar indicates the time window of significant differences between sampling conditions (interaction effect, see Results). **c)** Mean numerical distance effects by condition. *Left:* First half of samples in each choice trial. *Right:* Second half. **d)** Model RDM reflecting the sample values' extremity in terms of their absolute distance from the midpoint of the sample range (i.e., 5). **e)** Time course of extremity encoding in multivariate ERPs, plotted separately for each sampling condition. All error bars and shadings show SE.

Neurometric distortions

Recent studies of sequential number comparisons (without participant control over sampling) have shown that neural number representations can be distorted (e.g., compressed or anti-compressed) away from the perfectly linear distance structure of our idealized model RDMs (Figure 2.3*a*; see Materials and Methods). We used a "neurometric" approach (Spitzer et al., 2017) to test (i) whether

such distortions were replicated in our task and (ii) whether they differed between levels of control. To this end, we parameterized our model RDMs to reflect the distance structure of transformed values $v = sign(x + b) \times |x + b|^k$, where x are the numerical sample values (1-9 normalized to the range [-1, 1]), exponent k determines the shape of the transformation ($k < 1$ compression; $k = 1$ linear; $k > 1$ anti-compression), and b reflects a bias towards smaller ($b < 0$) or larger numbers ($b > 0$). Our ERP data, averaged across all conditions, were best explained by parameterizations $k > 1$ and $b > 0$ (Figure 2.4*a*; both p<0.003, t-tests of individual subject maxima against 1 and 0, respectively, averaged over parameterized distance and extremity). Thus, the neural number representation was anti-compressed and biased towards larger magnitudes (Figure 2.4*b*), strongly resembling the distortions observed in previous work (Luyckx et al., 2019; Spitzer et al., 2017). In comparisons between levels of control, however, we found no evidence for differences in the degree of anti-compression (Figure 2.4*c*, *left*; both p>0.545, t-tests of k against yoked baselines or bias; Figure 2.4*c*, *right*; both p>0.131, t-tests of b against yoked baselines). In other words, under full control, the encoding of numerical sample information was amplified (Figure 2.3*b* and *c*) without any notable changes in its general representational geometry.

Lastly, we examined whether our neurometric findings were also reflected in participants' sampling behavior. Stronger neural encoding of larger sample values (as suggested by the neurometric bias towards large numbers, cf. Figure 2.4*b*) may imply that these values (e.g., "9" or "8") may drive behavior more strongly than small numbers (e.g., "1" to "2"), despite their nominally identical diagnosticity for the options' mean values. If that was the case, participants in the full control condition should have been more likely to stop sampling after large numbers. Empirically, this should register in a relatively later mean position of large numbers (on average across trials) within the self-terminated sequences under full control, compared to the fixed-length sequences under partial control (where number values are expected to be uniformly distributed across the sequence by design). Indeed, we found that the mean relative position (relative to the sequence's length) of a sample increased with its numerical magnitude (1-9) in the full control conditions (p=0.005, linear trend analysis) but not in the partial control condition (p>0.78). In other words, participants showed a tendency to stop sampling after larger numbers, consistent with the finding of a neurometric bias towards larger numbers. We report this additional aspect of self-terminated sampling for the sake of completeness; our yoked design warrants that our findings about the effects of control over sampling are unaffected by it.

2.4 Discussion

Using variants of a numerical sampling paradigm and controlling for stimulus confounds, we observed increased choice accuracy when participants had control over the sampling process before committing to a choice. On the neural level, the behavioral benefit was reflected in a stronger encoding of the numerical sample information in multivariate EEG patterns and in a steeper build-up of centro-parietal choice signals. The key determinant of these effects was participants' control over "how much" information to sample. Freedom to decide only which options to sample, but not when to stop sampling, did not bring about the same effects, neither in behavior nor in neural signals.

Drawing on a well-established sequential sampling framework (Gold & Shadlen, 2007; O'Connell et al., 2012; Ratcliff & McKoon, 2008), our behavioral and neural findings provide a neurocognitive perspective on how control over sampling may boost choice accuracy. We observed no differences in early visual ERPs known to be modulated by top-down visual attention (Luck et al., 1994; Luck et al., 2000; Mangun & Hillyard, 1991), but a robust enhancement further downstream in

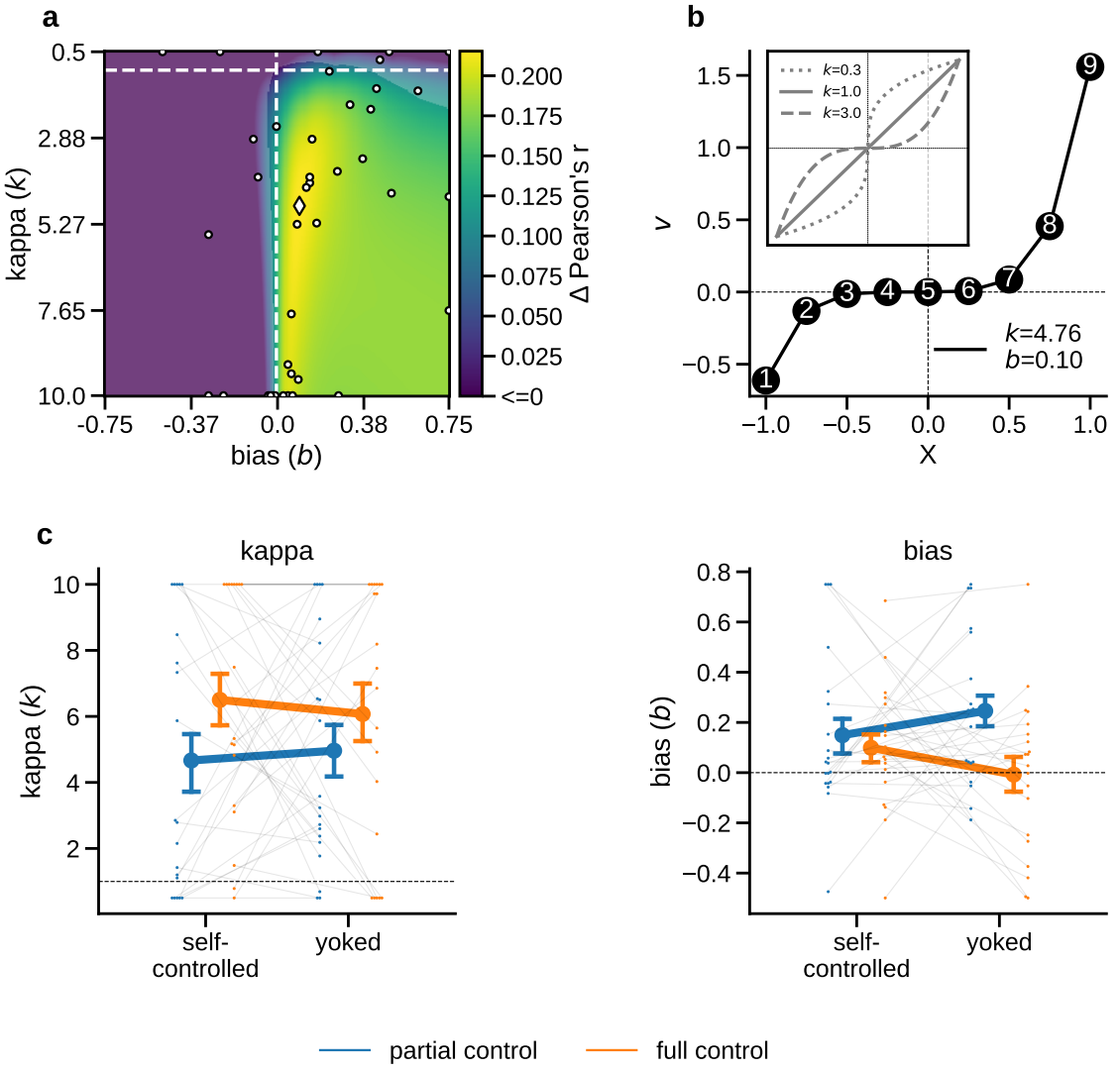

Figure 2.4. Neurometric distortions. **a)** Grand mean neurometric map, combined across all task conditions. Color scale indicates increase of encoding strength in multivariate ERPs (Δ r, averaged over distance- and extremity models) as a function of nonlinear distortions of numerical value ($k < 1$: compression; $k > 1$: anti-compression, b: bias). Dashed lines indicate linear ($k = 1$) and unbiased ($b = 0$) models. Parts of the map that are not overlaid with an opaque mask contain values with a significant increase relative to unbiased linear encoding ($p < 0.001$, corrected using FDR). White markers show maxima (diamond: mean; dots, individual participants). **b)** Neurometric function, parameterized according to the maximum mean correlation identified in *a*. Inset plots illustrate exemplary compressive ($k < 1$), linear ($k = 1$), and anti-compressive ($k > 1$) distortions. **c)** Neurometric parameter estimates in the individual sampling conditions, *left:* exponent (k); *right:* bias (b); see Materials and Methods and Results for details. Error bars show SE.

the processing hierarchy, at the level of symbolic number encoding (Ansari et al., 2005; Nieder & Dehaene, 2009). Our results replicate recent findings of a "neuronal numberline" in multivariate ERP patterns, where the neural representation of, for example, number "6" is more similar to that of "7" than to that of "9" (Luyckx et al., 2019; Sheahan et al., 2021; Spitzer et al., 2017; Teichmann et al., 2018). We found this multivariate encoding of numerical magnitude to be amplified under full control, mirroring the pattern observed in behavioral performance. Importantly, number encoding was already enhanced for samples occurring early in the trial, long before participants stopped sampling to make a final choice. Likewise, the behavioral benefit appeared driven by early and late samples alike, as indicated by the absence of differences in temporal weighting. Consistent

with these findings, we also observed a steeper rise in parietal indices of evidence accumulation (CPP/P3; O'Connell et al., 2012; Twomey et al., 2015) "across" samples, as if each individual sample contributed stronger evidence to the ongoing decision formation. In a sequential sampling framework where evidence is accumulated into a running decision variable (Glickman & Usher, 2019; Gold & Shadlen, 2007; Kiani et al., 2008; O'Connell et al., 2012; Ratcliff & McKoon, 2008), our EEG and behavioral findings may thus both be attributable to an improvement in numerical evidence processing.

One possible explanation for our findings relates to motivational factors. Previous work has shown that the ability to actively control the environment and/or one's subjective experiences can have beneficial effects, for example, on memory (Murty et al., 2015; Voss et al., 2011), self-regulation, and error monitoring (Legault & Inzlicht, 2013), learning and inductive inference (Gureckis & Markant, 2012; Markant & Gureckis, 2014), and various other aspects of cognition and behavior (Leotti & Delgado, 2011; Leotti et al., 2010; Murayama et al., 2016; Patall, 2012; Patall et al., 2008). Our findings add to these literature by showing that control can also confer benefits in sample-based decision-making, specifically when participants can control when to stop sampling. While the extrinsic rewards for choice accuracy were identical across our task conditions, control over stopping can add an incentive to optimize the time spent on a trial (Ostwald et al., 2015; Tickle et al., 2020). There is typically a trade-off between speed and accuracy of task execution (Heitz, 2014), such that faster decisions come at the cost of lower accuracy (but see Gigerenzer et al., 2011), However, the present findings under full control cannot be explained by such a trade-off, given that we observed benefits relative to yoked trials of identical length. As we used exact copies of participant-generated sampling sequences in our baseline conditions, we can also rule out the possibility that the results are attributable to amplification effects (Hertwig & Pleskac, 2010), where participants tend to stop sampling when the cumulative difference between options happens to be large (leading to objectively easier trials; see below). With these simpler explanations ruled out, our findings suggest that control per se may lead to more efficient sample encoding, potentially through increased task engagement when decision time can be optimized on a trial-by-trial basis.

We found no differences between conditions in the temporal weighting of sample information over the course of a trial. A clear recency effect (relative overweighting of late samples) was evident in all task conditions, including yoked baselines. This pattern appears to be at odds with a previous meta-analysis of numerical sampling studies (Wulff et al., 2018), where recency effects were observed solely in conditions with full agency over sampling. However, the present findings are consistent with routine observations of recency effects in other sequential integration tasks where sample presentation is entirely experimenter-controlled (Anderson, 1964; Cheadle et al., 2014; Kang & Spitzer, 2021; Luyckx et al., 2019; Spitzer et al., 2017; Tsetsos et al., 2012; D. J. Weiss & Anderson, 1969; Wyart et al., 2015). Here, using carefully designed yoked control conditions, we found no evidence that the strength of recency effects (which may arise, e.g., by forgetting, or leakage of sample information over time; Usher & McClelland, 2001) would depend on the level of control over sampling. We also found no differences in the representational geometry of the sampled information in neural signals. Neurometric analysis showed an anti-compression of numerical values (Luyckx et al., 2019; Spitzer et al., 2017) in all conditions, regardless of the level of control. The absence of differences in these more qualitative aspects of information processing in our tasks suggests that the cognitive benefits of full control may best be described as an overall increase in the gain of neural processing (Donner & Nieuwenhuis, 2013; Eldar et al., 2013; Murphy et al., 2016), which may amplify the critical decisional information in a sample (here, numerical magnitude).

None of the benefits observed under full control were evident in the partial control condition, where participants could only decide which option to sample next, but not when to terminate sampling. Although this condition gave participants some level of agency (relative to the yoked conditions without control; Chambon et al., 2020; A. Weiss et al., 2021), we suspect that it may not have induced a strong sense of control over the task. It even seems possible that participants may have perceived the requirement to perform a prescribed number of sampling actions as externally controlled and a cognitive burden (see also Sullivan-Toole et al., 2017). Indeed, post hoc examination of left/right sampling patterns showed that our participants resorted to stereotypical sampling routines (either alternating between options: "a-b-a-b-..." or sampling first one option and then the other: "a-a-a-...-b-b-b") in 67.61% of trials (relative to the yoked conditions without control; for related findings, see Hills & Hertwig, 2010). In other words, participants made little use of the freedom to vary their left/right sampling strategy trial by trial (and/or sample by sample), potentially due to a lack of perceived benefits (Dixon & Christoff, 2012). In this light, it is perhaps not surprising that we found no processing enhancements under partial control, in either behavior or neural signals.

Numerical sampling tasks similar to ours have been used extensively in the past to study decisions from experience (Hertwig et al., 2004) in complement to the common use of symbolic descriptions to study risky choice (Juechems et al., 2021; Kahneman & Tversky, 1979). Experience-based choices can differ systematically from description-based choice, especially in terms of probability weighting (Hertwig & Erev, 2009; Wulff et al., 2018). A much-discussed aspect of this "description-experience gap" is that participants in experience-based tasks tend to rely on relatively few samples (Hau et al., 2010; Plonsky et al., 2015; Wulff et al., 2018). Also in our experiment, participants in the full control condition chose to sample less than they could have (Furl & Averbeck, 2011). Although one explanation is that small samples can render choices objectively simpler (Hertwig & Pleskac, 2008, 2010), our findings suggest that small samples may also defy typical accuracy trade-offs if the decision to stop sampling lies in the autonomy of the sampling agent (see also Petitet et al., 2021). Granting participants' full control over sampling may thus not only enable but directly promote reliance on small samples through more efficient processing of the sample evidence.

Finally, our multivariate EEG analysis also showed a neural signature of the samples' "extremity" (Figure 2.3*d* and *e*), which did not differ between levels of control over sampling. Future work may investigate the potential significance of this finding with respect to the role of extreme events in experience-based decisions (e.g., Ludvig et al., 2018; Ludvig et al., 2014; Ludvig & Spetch, 2011).

In summary, we found that control over sampling can enhance the neural encoding of decision information and improve choice accuracy. The results add to a growing collection of findings that exercising agency can benefit performance in cognitive tasks and shed light on the neural processes that may support such benefits.

2.5 Acknowledgments

We thank Agnessa Karapetian, Clara Wicharz, Jann Wäscher, Yoonsang Lee, and Zhiqi Kang for help with data collection, Dirk Ostwald and Casper Kerrén for helpful discussions and feedback, and Susannah Goss for editorial assistance.

2.6 Funding

BS is supported by a European Research Council Consolidator Grant ERC-2020-COG-101000972.

Conflict of interest statement. The authors declare no competing interests.

2.7 Supplemental material

The dataset in BIDS format is available on GIN (https://doi.org/10.12751/g-node.dtyh14).

All code is available on Zenodo (Analysis code: https://doi.org/10.5281/zenodo.5929222; experiment code: https://doi.org/10.5281/zenodo.3354368).

The preprint for this chapter is available on BioRxiv under: https://doi.org/10.1101/2021.06.03.446960

2.8 Ethics information

The study was approved by the ethics committee of the Max Planck Institute for Human Development.

2.9 Author contributions

SA, RH, and BS: Conceptualization, project administration, writing - review and editing.

SA and BS: Methodology, writing - original draft.

SA: Formal analysis, investigation, validation, visualization, data curation, software.

BS: Supervision.

RH: Resources.

Corresponding author: Stefan Appelhoff

For a more detailed overview of author contributions, please see Appendix B.

2.10 References

Anderson, N. H. (1964). Test of a model for number-averaging behavior [Place: US Publisher: Psychonomic Society]. *Psychonomic Science*, *1*(7), 191–192. https://doi.org/10.3758/BF03342858

Ansari, D., Garcia, N., Lucas, E., Hamon, K., & Dhital, B. (2005). Neural correlates of symbolic number processing in children and adults. *NeuroReport*, *16*(16), 1769–1773. https://doi.org/10.1097/01.wnr.0000183905.23396.f1

Appelhoff, S., Sanderson, M., Brooks, T. L., Vliet, M. v., Quentin, R., Holdgraf, C., Chaumon, M., Mikulan, E., Tavabi, K., Höchenberger, R., Welke, D., Brunner, C., Rockhill, A. P., Larson, E., Gramfort, A., & Jas, M. (2019). MNE-BIDS: Organizing electrophysiological data into the BIDS format and facilitating their analysis. *Journal of Open Source Software*, *4*(44), 1896. https://doi.org/10.21105/joss.01896

Chambon, V., Théro, H., Vidal, M., Vandendriessche, H., Haggard, P., & Palminteri, S. (2020). Information about action outcomes differentially affects learning from self-determined versus imposed choices. *Nature Human Behaviour*, *4*(10), 1067–1079. https://doi.org/10.1038/s41562-020-0919-5

Cheadle, S., Wyart, V., Tsetsos, K., Myers, N., de Gardelle, V., Herce Castañón, S., & Summerfield, C. (2014). Adaptive Gain Control during Human Perceptual Choice. *Neuron*, *81*(6), 1429–1441. https://doi.org/10.1016/j.neuron.2014.01.020

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making*, *7*(1), 25–47.

Dixon, M. L., & Christoff, K. (2012). The Decision to Engage Cognitive Control Is Driven by Expected Reward-Value: Neural and Behavioral Evidence (F. P. de Lange, Ed.). *PLoS ONE*, *7*(12), e51637. https://doi.org/10.1371/journal.pone.0051637

Donner, T. H., & Nieuwenhuis, S. (2013). Brain-wide gain modulation: The rich get richer. *Nature Neuroscience*, *16*(8), 989–990. https://doi.org/10.1038/nn.3471

Eimer, M. (1998). The lateralized readiness potential as an on-line measure of central response activation processes. *Behavior Research Methods, Instruments, & Computers*, *30*(1), 146–156. https://doi.org/10.3758/BF03209424

Eldar, E., Cohen, J. D., & Niv, Y. (2013). The effects of neural gain on attention and learning [Number: 8 Publisher: Nature Publishing Group]. *Nature Neuroscience*, *16*(8), 1146–1153. https://doi.org/10.1038/nn.3428

Fleischhut, N., Artinger, F., Olschewski, S., Volz, K., & Hertwig, R. (2014). Sampling of Social Information: Decisions from Experience in Bargaining. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 1048–1053). Cognitive Science Society.

Flowerday, T., & Schraw, G. (2003). Effect of Choice on Cognitive and Affective Engagement [Publisher: Routledge _eprint: https://doi.org/10.1080/00220670309598810]. *The Journal of Educational Research*, *96*(4), 207–215. https://doi.org/10.1080/00220670309598810

Flowerday, T., Schraw, G., & Stevens, J. (2004). The Role of Choice and Interest in Reader Engagement [Publisher: Taylor & Francis, Ltd.]. *The Journal of Experimental Education*, *72*(2), 93–114. Retrieved April 21, 2021, from https://www.jstor.org/stable/20157361

Furl, N., & Averbeck, B. B. (2011). Parietal Cortex and Insula Relate to Evidence Seeking Relevant to Reward-Related Decisions. *Journal of Neuroscience*, *31*(48), 17572–17582. https://doi.org/10.1523/JNEUROSCI.4236-11.2011

Gigerenzer, G., Hertwig, R., & Pachur, T. (2011). *Heuristics*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199744282.001.0001

Glickman, M., & Usher, M. (2019). Integration to boundary in decisions between numerical sequences. *Cognition*, *193*, 104022. https://doi.org/10.1016/j.cognition.2019.104022

Gold, J. I., & Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience*, *30*(1), 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Gonzalez, C., & Mehlhorn, K. (2016). Framing From Experience: Cognitive Processes and Predictions of Risky Choice. *Cognitive Science*, *40*(5), 1163–1191. https://doi.org/10.1111/cogs.12268

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing

outputs of neuroimaging experiments [Number: 1 Publisher: Nature Publishing Group]. *Scientific Data*, *3*(1), 160044. https://doi.org/10.1038/sdata.2016.44

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python [Publisher: Frontiers]. *Frontiers in Neuroscience*, *7*(267). https://doi.org/10.3389/fnins.2013.00267

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2016). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data [Publisher: MIT Press]. *Journal of Cognitive Neuroscience*, *29*(4), 677–697. https://doi.org/10.1162/jocn_a_01068

Gureckis, T. M., & Markant, D. B. (2012). Self-Directed Learning: A Cognitive and Computational Perspective. *Perspectives on Psychological Science*, *7*(5), 464–481. https://doi.org/10.1177/1745691612454304

Hau, R., Pleskac, T. J., & Hertwig, R. (2010). Decisions from experience and statistical probabilities: Why they trigger different choices than a priori probabilities. *Journal of Behavioral Decision Making*, *23*(1), 48–68. https://doi.org/10.1002/bdm.665

Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description-experience gap in risky choice: The role of sample size and experienced probabilities. *Journal of Behavioral Decision Making*, *21*(5), 493–518. https://doi.org/10.1002/bdm.598

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*, *8*(150). https://doi.org/10.3389/fnins.2014.00150

Hertwig, R. (2015). Decisions from Experience. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell Handbook of Judgment and Decision Making* (pp. 239–267). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118468333.ch8

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from Experience and the Effect of Rare Events in Risky Choice. *Psychological Science*, *15*(8), 534–539. https://doi.org/10.1111/j.0956-7976.2004.00715.x

Hertwig, R., & Erev, I. (2009). The descriptionexperience gap in risky choice. *Trends in Cognitive Sciences*, *13*(12), 517–523. https://doi.org/10.1016/j.tics.2009.09.004

Hertwig, R., & Pleskac, T. J. (2008). The game of life: How small samples render choice simpler. In N. Chater & M. Oaksford (Eds.), *The Probabilistic Mind: Prospects for Bayesian cognitive science* (pp. 209–236). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199216093.003.0010

Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: Why small samples? *Cognition*, *115*(2), 225–237. https://doi.org/10.1016/j.cognition.2009.12.009

Hills, T. T., & Hertwig, R. (2010). Information Search in Decisions From Experience: Do Our Patterns of Sampling Foreshadow Our Decisions? *Psychological Science*, *21*(12), 1787–1792. https://doi.org/10.1177/0956797610387443

Juechems, K., Balaguer, J., Spitzer, B., & Summerfield, C. (2021). Optimal utility and probability functions for agents with finite computational precision [Publisher: National Academy of Sciences Section: Social Sciences]. *Proceedings of the National Academy of Sciences*, *118*(2). https://doi.org/10.1073/pnas.2002232118

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–292. https://doi.org/10.2307/1914185

Kang, Z., & Spitzer, B. (2021). Concurrent visual working memory bias in sequential integration of approximate number. *Scientific Reports*, *11*(1), 5348. https://doi.org/10.1038/s41598-021-84232-7

Katz, I., & Assor, A. (2007). When Choice Motivates and When It Does Not. *Educational Psychology Review*, *19*(4), 429–442. https://doi.org/10.1007/s10648-006-9027-y

Kiani, R., Hanks, T. D., & Shadlen, M. N. (2008). Bounded Integration in Parietal Cortex Underlies Decisions Even When Viewing Duration Is Dictated by the Environment [Publisher: Society for Neuroscience Section: Articles]. *Journal of Neuroscience*, *28*(12), 3017–3029. https://doi.org/10.1523/JNEUROSCI.4761-07.2008

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412. https://doi.org/10.1016/j.tics.2013.06.007

Legault, L., & Inzlicht, M. (2013). Self-determination, self-regulation, and the brain: Autonomy improves performance by enhancing neuroaffective responsiveness to self-regulation failure. *Journal of Personality and Social Psychology*, *105*(1), 123–138. https://doi.org/10.1037/a0030426

Leotti, L. A., & Delgado, M. R. (2011). The Inherent Reward of Choice [Publisher: SAGE Publications Inc]. *Psychological Science*, *22*(10), 1310–1318. https://doi.org/10.1177/0956797611417005

Leotti, L. A., Iyengar, S. S., & Ochsner, K. N. (2010). Born to choose: The origins and value of the need for control. *Trends in Cognitive Sciences*, *14*(10), 457–463. https://doi.org/10.1016/j.tics.2010.08.001

Luck, S. J., Hillyard, S. A., Mouloua, M., Woldorff, M. G., Clark, V. P., & Hawkins, H. L. (1994). Effects of spatial cuing on luminance detectability: Psychophysical and electrophysiological evidence for early selection. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(4), 887–904. https://doi.org/10.1037/0096-1523.20.4.887

Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, *4*(11), 432–440. https://doi.org/10.1016/S1364-6613(00)01545-X

Ludvig, E. A., Madan, C. R., McMillan, N., Xu, Y., & Spetch, M. L. (2018). Living near the edge: How extreme outcomes and their neighbors drive risky choice. *Journal of Experimental Psychology: General*, *147*(12), 1905–1918. https://doi.org/10.1037/xge0000414

Ludvig, E. A., Madan, C. R., & Spetch, M. L. (2014). Extreme Outcomes Sway Risky Decisions from Experience: Risky Decisions and Extreme Outcomes. *Journal of Behavioral Decision Making*, *27*(2), 146–156. https://doi.org/10.1002/bdm.1792

Ludvig, E. A., & Spetch, M. L. (2011). Of Black Swans and Tossed Coins: Is the Description-Experience Gap in Risky Choice Limited to Rare Events? (A. Sirigu, Ed.). *PLoS ONE*, *6*(6), e20262. https://doi.org/10.1371/journal.pone.0020262

Luyckx, F., Nili, H., Spitzer, B., & Summerfield, C. (2019). Neural structure mapping in human probabilistic reward learning (D. Lee, J. I. Gold, D. Lee, & M. Chafee, Eds.) [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *8*, e42816. https://doi.org/10.7554/eLife.42816

Mangun, G. R., & Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(4), 1057–1074. https://doi.org/10.1037/0096-1523.17.4.1057

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Markant, D. B., & Gureckis, T. M. (2014). Is it better to select or to receive? Learning via active and passive hypothesis testing. *Journal of Experimental Psychology. General*, *143*(1), 94–122. https://doi.org/10.1037/a0032108

Murayama, K., Izuma, K., Aoki, R., & Matsumoto, K. (2016). Your Choice Motivates You in the Brain: The Emergence of Autonomy Neuroscience. *Recent Developments in Neuroscience Research on Human Motivation* (pp. 95–125). Emerald Group. https://doi.org/10.1108/S0749-742320160000019004

Murayama, K., Matsumoto, M., Izuma, K., Sugiura, A., Ryan, R. M., Deci, E. L., & Matsumoto, K. (2015). How Self-Determined Choice Facilitates Performance: A Key Role of the Ventromedial Prefrontal Cortex. *Cerebral Cortex*, *25*(5), 1241–1251. https://doi.org/10.1093/cercor/bht317

Murphy, P. R., Boonstra, E., & Nieuwenhuis, S. (2016). Global gain modulation generates time-dependent urgency during perceptual choice in humans. *Nature Communications*, *7*(1), 13526. https://doi.org/10.1038/ncomms13526

Murty, V. P., DuBrow, S., & Davachi, L. (2015). The Simple Act of Choosing Influences Declarative Memory [Publisher: Society for Neuroscience Section: Articles]. *Journal of Neuroscience*, *35*(16), 6255–6264. https://doi.org/10.1523/JNEUROSCI.4181-14.2015

Neri, P., Parker, A. J., & Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation [Number: 6754 Publisher: Nature Publishing Group]. *Nature*, *401*(6754), 695–698. https://doi.org/10.1038/44409

Nieder, A. (2016). The neuronal code for number [Bandiera_abtest: a Cg_type: Nature Research Journals Number: 6 Primary_atype: Reviews Publisher: Nature Publishing Group Subject_term: Cognitive neuroscience;Decision;Working memory Subject_term_id: cognitive-neuroscience;decision;working-memory]. *Nature Reviews Neuroscience*, *17*(6), 366–382. https://doi.org/10.1038/nrn.2016.40

Nieder, A., & Dehaene, S. (2009). Representation of Number in the Brain. *Annual Review of Neuroscience*, *32*(1), 185–208. https://doi.org/10.1146/annurev.neuro.051508.135550

Nolan, H., Whelan, R., & Reilly, R. B. (2010). FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection. *Journal of Neuroscience Methods*, *192*(1), 152–162. https://doi.org/10.1016/j.jneumeth.2010.07.015

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans [Number: 12 Publisher: Nature Publishing Group]. *Nature Neuroscience*, *15*(12), 1729–1735. https://doi.org/10.1038/nn.3248

Ostwald, D., Starke, L., & Hertwig, R. (2015). A normative inference approach for optimal sample sizes in decisions from experience. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01342

Patall, E. A. (2012). The Motivational Complexity of Choosing: A Review of Theory and Research. In R. M. Ryan (Ed.), *The Oxford Handbook of Human Motivation* (pp. 247–279). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399820.013.0015

Patall, E. A., Cooper, H., & Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: A meta-analysis of research findings. *Psychological Bulletin*, *134*(2), 270–300. https://doi.org/10.1037/0033-2909.134.2.270

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pernet, C. R., Appelhoff, S., Gorgolewski, K. J., Flandin, G., Phillips, C., Delorme, A., & Oostenveld, R. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography [Number: 1 Publisher: Nature Publishing Group]. *Scientific Data*, *6*(1), 103. https://doi.org/10.1038/s41597-019-0104-8

Petitet, P., Attaallah, B., Manohar, S. G., & Husain, M. (2021). The computational cost of active information sampling before decision-making under uncertainty. *Nature Human Behaviour*, *5*, 935–946. https://doi.org/10.1038/s41562-021-01116-6

Pisauro, M. A., Fouragnan, E., Retzler, C., & Philiastides, M. G. (2017). Neural correlates of evidence accumulation during value-based decisions revealed via simultaneous EEG-fMRI. *Nature Communications*, *8*(1), 15808. https://doi.org/10.1038/ncomms15808

Plonsky, O., Teodorescu, K., & Erev, I. (2015). Reliance on small samples, the wavy recency effect, and similarity-based learning. *Psychological Review*, *122*(4), 621–647. https://doi.org/10.1037/a0039413

Polich, J. (2007). Updating P300: An Integrative Theory of P3a and P3b. *Clinical neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Rakow, T., Demes, K. A., & Newell, B. R. (2008). Biased samples not mode of presentation: Re-examining the apparent underweighting of rare events in experience-based choice. *Organizational Behavior and Human Decision Processes*, *106*(2), 168–179. https://doi.org/10.1016/j.obhdp.2008.02.001

Rakow, T., & Newell, B. R. (2010). Degrees of uncertainty: An overview and framework for future research on experience-based choice. *Journal of Behavioral Decision Making*, *23*(1), 1–14. https://doi.org/10.1002/bdm.681

Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Sheahan, H., Luyckx, F., Nelli, S., Teupe, C., & Summerfield, C. (2021). Neural state space alignment for magnitude generalization in humans and recurrent networks. *Neuron*, *109*, 1214–1226. https://doi.org/10.1016/j.neuron.2021.02.004

Spitzer, B., Blankenburg, F., & Summerfield, C. (2016). Rhythmic gain control during supramodal integration of approximate number. *NeuroImage*, *129*, 470–479. https://doi.org/10.1016/j.neuroimage.2015.12.024

Spitzer, B., Waschke, L., & Summerfield, C. (2017). Selective overweighting of larger magnitudes during noisy numerical comparison. *Nature Human Behaviour*, *1*(8), 1–8. https://doi.org/10.1038/s41562-017-0145

Sullivan-Toole, H., Richey, J. A., & Tricomi, E. (2017). Control and Effort Costs Influence the Motivational Consequences of Choice. *Frontiers in Psychology*, *8*(675). https://doi.org/10.3389/fpsyg.2017.00675

Teichmann, L., Grootswagers, T., Carlson, T., & Rich, A. N. (2018). Decoding Digits and Dice with Magnetoencephalography: Evidence for a Shared Representation of Magnitude [Publisher: MIT Press]. *Journal of Cognitive Neuroscience*, *30*(7), 999–1010. https://doi.org/10.1162/jocn_a_01257

Thaler, L., Schütz, A. C., Goodale, M. A., & Gegenfurtner, K. R. (2013). What is the best fixation target? The effect of target shape on stability of fixational eye movements. *Vision Research*, *76*, 31–42. https://doi.org/10.1016/j.visres.2012.10.012

Tickle, H., Tsetsos, K., Speekenbrink, M., & Summerfield, C. (2020). *Optional Stopping in a Heteroscedastic World* (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/t7dn2

Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal [ISBN: 9781119569107 Publisher: National Academy of Sciences Section: Biological Sciences]. *Proceedings of the National Academy of Sciences*, *109*(24), 9659–9664. https://doi.org/10.1073/pnas.1119569109

Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience*, *42*(1), 1636–1643. https://doi.org/10.1111/ejn.12936

Ungemach, C., Chater, N., & Stewart, N. (2009). Are Probabilities Overweighted or Underweighted When Rare Outcomes Are Experienced (Rarely)? *Psychological Science*, *20*(4), 473–479. https://doi.org/10.1111/j.1467-9280.2009.02319.x

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592. https://doi.org/10.1037/0033-295X.108.3.550

Voss, J. L., Gonsalves, B. D., Federmeier, K. D., Tranel, D., & Cohen, N. J. (2011). Hippocampal brain-network coordination during volitional exploratory behavior enhances learning. *Nature Neuroscience*, *14*(1), 115–120. https://doi.org/10.1038/nn.2693

Weiss, A., Chambon, V., Lee, J. K., Drugowitsch, J., & Wyart, V. (2021). Interacting with volatile environments stabilizes hidden-state inference and its brain signatures [Number: 1 Publisher: Nature Publishing Group]. *Nature Communications*, *12*(1), 2228. https://doi.org/10.1038/s41467-021-22396-6

Weiss, D. J., & Anderson, N. H. (1969). Subjective averaging of length with serial presentation. *Journal of Experimental Psychology*, *82*(1, Pt.1), 52–63. https://doi.org/10.1037/h0028028

Winkler, I., Debener, S., Müller, K., & Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP [ISSN: 1558-4615]. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 4101–4105. https://doi.org/10.1109/EMBC.2015.7319296

Wulff, D., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*, *144*(2), 140–176. https://doi.org/10.1037/bul0000115

Wyart, V., Myers, N. E., & Summerfield, C. (2015). Neural Mechanisms of Human Perceptual Choice Under Focused and Divided Attention [Publisher: Society for Neuroscience Section: Articles]. *Journal of Neuroscience*, *35*(8), 3485–3498. https://doi.org/10.1523/JNEUROSCI.3276-14.2015

3 | EEG-representational geometries of psychometric distortions in approximate numerical judgment

Stefan Appelhoff, Ralph Hertwig & Bernhard Spitzer

The contents of this chapter have been submitted for publication in *PLOS Computational Biology*.

Appelhoff, **S.**, Hertwig, R. & Spitzer, B. (2022). EEG-representational geometries of psychometric distortions in approximate numerical judgment. (submitted to *PLOS Computational Biology*)

Abstract

When judging the average value of sample stimuli (e.g., numbers) people tend to either over- or underweight extreme sample values, depending on task context. In a context of overweighting, recent work has shown that extreme sample values were overly represented also in neural signals, in terms of an anti-compressed geometry of number samples in multivariate electroencephalography (EEG) patterns. Here, we asked whether neural representational geometries may also reflect underweighting of extreme values (i.e., compression) which has been observed behaviorally in a great variety of tasks. We used a simple experimental manipulation (instructions to average a single-stream or to compare dual-streams of samples) to induce compression or anti-compression in behavior when participants judged rapid number sequences. Model-based representational similarity analysis (RSA) replicated the previous finding of neural anti-compression in the dual-stream task, but failed to provide evidence for neural compression in the single-stream task, despite the evidence for compression in behavior. Instead, the results suggested enhanced neural processing of extreme values in either task, regardless of whether extremes were over- or underweighted in subsequent behavioral choice. We further observed more general differences in the neural representation of the sample information between the two tasks. The results suggest enhanced processing of extreme values as the brain's default. Such a default raises new questions about the origin of common psychometric distortions, such as diminishing sensitivity for larger values.

3.1 Introduction

When making decisions about magnitudes such as numbers, people tend to distort sample information away from its true value. A commonly observed distortion is a compression of magnitude, where extreme (or outlying) samples receive relatively less weight than prescribed by a linear and, according to common interpretation, normative transformation of objective into psychological or subjective values (Bernoulli, 1954; Fechner, 1860; Juechems et al., 2021; Li et al., 2017; Tversky & Kahneman, 1992; Vandormael et al., 2017). However, in some task contexts, the opposite type of distortion has been observed, that is, an anti-compression, where extreme or outlying samples are overweighted (Clarmann von Clarenau et al., 2022; Kunar et al., 2017; Luyckx et al., 2019; Spitzer et al., 2017; Tsetsos et al., 2012; Vanunu et al., 2020).

Using electroencephalographic (EEG) recordings and multivariate representational similarity analysis (RSA), recent work has identified a potential neural signature of such psychometric distortions. During processing of symbolic number samples, multivariate EEG patterns have been characterized by a "numerical distance effect", where the representational similarity of, for instance, "4" and "5" is larger than that between "4" and "6", which in turn is larger than that between "3" and "7", and so forth (Appelhoff et al., 2022; Luyckx et al., 2019; Sheahan et al., 2021; Spitzer et al., 2017; Teichmann et al., 2018). Intriguingly, in a multi-sample decision task that promoted anti-compression of number samples in behavior, the "neural numberline" underlying the numerical distance effect was found to be anti-compressed as well (Appelhoff et al., 2022; Luyckx et al., 2019; Spitzer & Haegens, 2017). The findings suggested that behaviorally relevant distortions in multi-sample decisions may occur already when the individual samples are being processed.

However, such "neurometric" signature of psychometric distortion has thus far only been reported in tasks that promoted anti-compression, that is, a selective overweighting of extreme (outlying) sample values in behavior. A much more common observation in other task contexts is a compression of magnitude, where extreme values are underweighted, for instance, in psychophysical tasks (Fechner, 1860; Stevens, 1957; Wyart et al., 2012), in studies of numerical cognition (Dehaene, 2003; Longo & Lourenco, 2007; Nieder & Dehaene, 2009; Nieder & Miller, 2003), and in behavioral economics experiments (Kellen et al., 2016; McAllister & Tarbert, 1999; Tversky & Kahneman, 1992). To what extent large-scale neural patterns, as recorded with human EEG, may also reflect psychometric compression is still unknown.

In the present study, we capitalized on recent progress in understanding the experimental factors that may mediate whether people compress or anti-compress magnitudes in decision making (see also Summerfield & Parpart, 2022). Specifically, in a recent behavioral study, we found compression when judging the average of a single sequence, but anti-compression when comparing dual-streams of number samples (Clarmann von Clarenau et al., 2022). Here, we adopted this experimental manipulation to examine the neural signatures of compressive (as compared to anti-compressive) number processing in multivariate EEG patterns.

Behaviorally, the results confirmed a compression of numerical values in the single-stream task, and an anti-compression of the same values in the dual-stream task. In neural signals, we replicated the finding of an anti-compressed number representation in the dual-stream task. Surprisingly, however, we found no evidence for neural compression in the single-stream task. Instead, we observed more general differences in the neural representation of the sample information. Whereas in the dual-stream task, the samples' neural geometry predominantly reflected their abstract magnitude, the single-stream task was associated with a more direct, non-quantitative representation of

the concrete sample stimuli. The results relativize the diagnosticity of sample-level EEG-metrics for psychometric distortions. They also suggest a default mode of processing, namely, enhanced neural processing of extreme sample values, regardless of whether they are over- or underweighted in subsequent behavior.

3.2 Results

3.2.1 Behavioral results

As expected, mean choice accuracy (Figure 3.1*b*) was higher in the single-stream task (80.5% ± 0.8% SE) than in the dual-stream task [76.3% ± 0.7% SE; t(29)=5.51, p<0.001, d=1.02; paired t-test]. This suggests that comparing two streams was more difficult than averaging a single-stream (of otherwise physically identical inputs; Figure 3.1*a*).

To characterize how the numerical value (1-9) of a sample influenced subsequent choice, we calculated model-free decision weights (see Materials and Methods). Descriptively, the weighting curve showed a concave shape (indicating compression) in the single-stream task (Figure 3.1*c*, *left*), whereas an convex shape (indicating anti-compression) was evident in the dual-stream task (Figure 3.1*c*, *right*).

For quantitative analysis, we fitted a psychometric model (see Materials and Methods), which characterizes the transformation of sample values (1-9) as a sign-preserving power function with exponent kappa (k; where $k < 1$ indicates compression and $k > 1$ anti-compression). The model further includes parameters for overall bias (b) towards larger or smaller numbers ($b >$ or < 0) and decision noise (s, see Materials and Methods).

The best-fitting parameter estimates are shown in Figure 3.1*d*. Of main interest was parameter kappa (k), which indicates the extent to which a weighting policy is compressed or anti-compressed relative to a linear weighting ($k = 1$). Indeed, k was significantly smaller than 1 in the single-stream task [M=0.82, SE=0.07, t(29)=-2.43, p=0.02, d=0.44; one-sample t-test against 1] and significantly larger than 1 in the dual-stream task [M=1.82, SE=0.16, t(29)=5.21, p<0.001, d=0.95], which confirms robust compression in single-stream averaging, and robust anti-compression in dual-stream comparison. The essential difference in transformation (Equation 3.1) implied by these k-values is illustrated in Figure 3.1*c* (insets). While the transformation is concave (i.e., shallower towards the extremes) in the single-stream task, it is convex (i.e., steeper towards the extremes) in the dual-stream task.

Examining bias (b), we found significantly positive values both in the single-stream task [M=0.02, SE=0.01, t(29)=3.52, p=0.001, d=0.64] and in the dual-stream task [M=0.21, SE=0.06, t(29)=3.34, p=0.002, d=0.61]. Thus, judgments in both tasks were overall biased towards larger numbers, which is consistent with previous work (Clarmann von Clarenau et al., 2022; Luyckx et al., 2019; Spitzer et al., 2017). Finally, noise (s) was significantly higher in the dual-stream (M=1.23, SE=0.07) than in the single-stream task [M=0.98, SE=0.06; t(29)=-4.40, p<0.001, d=0.70; paired t-test]. This is consistent with the lower level of accuracy in the dual-stream task (see Figure 3.1*b*).

Together, our experimental manipulation was successful in inducing opposite types of psychometric distortions, with identical stimulus inputs in the two task conditions. Whereas decision weighting was compressed (suggesting relative underweighting of extreme values) in the single-stream task, it was anti-compressed (suggesting relative overweighting of extreme values) in the dual-stream task.

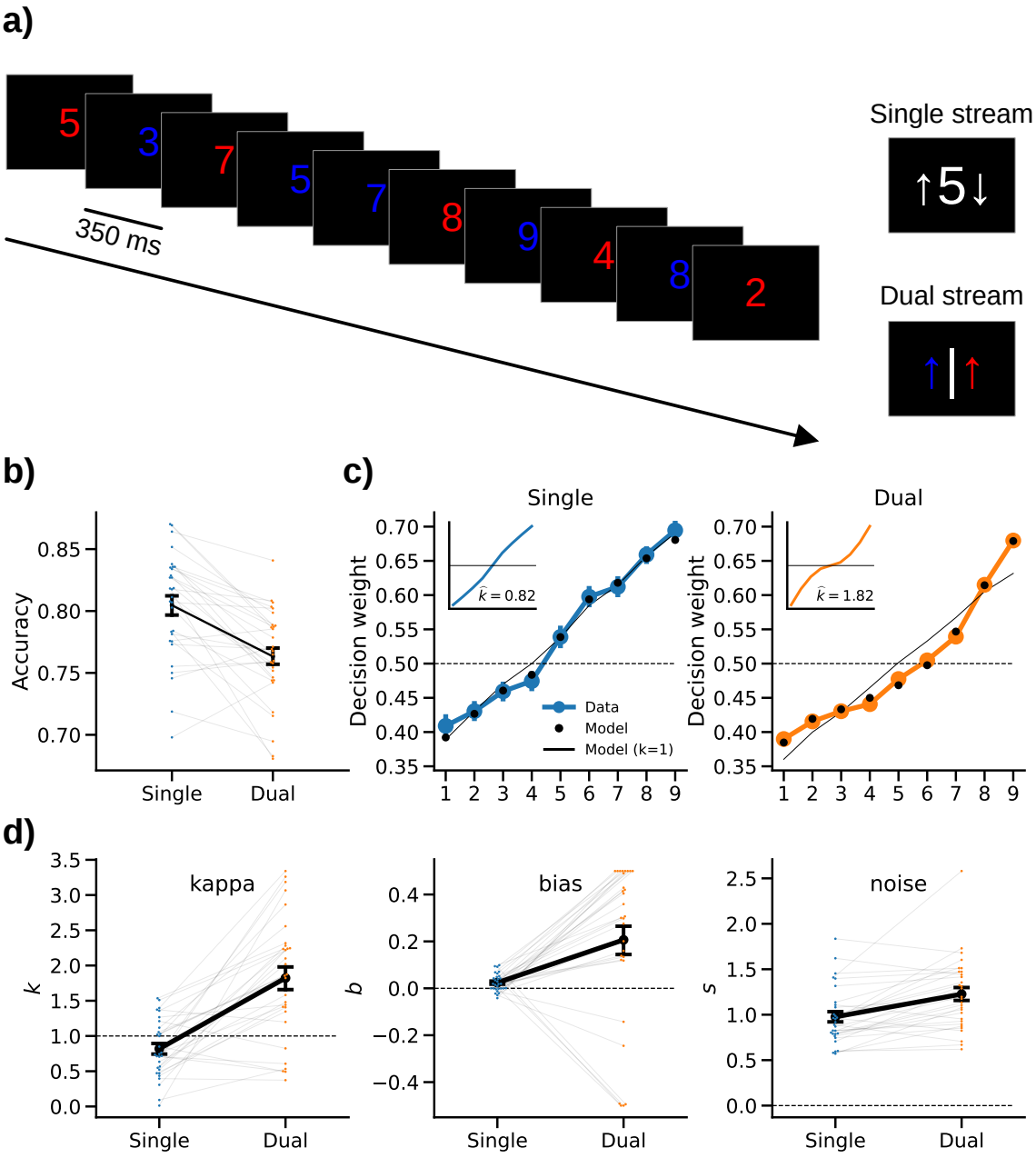

Figure 3.1. Experimental paradigm and behavioral results. a) Each participant performed two variants of a sequential number integration task. *Left:* In both variants, participants viewed a stream of ten rapidly presented digits (five in red and five in blue color; in random serial order) drawn from 1 to 9 (uniform random). In the single-stream averaging task, participants were asked to judge whether the average value of all ten samples was higher or lower than 5 (ignoring the samples' colors). In the dual-stream comparison task, participants were asked to report whether the red or the blue samples had the higher average value. *Right:* In both tasks, the response mapping onto left/right button presses was randomized across trials and cued only after sample presentation, in order to avoid motor preparation confounds. **b)** Mean accuracy (proportion correct choices) in the two tasks. **c)** Decision weights of number values in the single-stream (*left*) and dual-stream (*right*) tasks. Inset plots illustrate the shape of distortion implied by the best-fitting k (see panel d) according to model Equation 3.1, with b set to 0 for visual comparison between task conditions. **d)** Parameter estimates from fitting our psychometric model to the empirical choice data (cf. c). Error bars in all panels show SE.

3.2.2 EEG results

Multivariate (RSA) results

Turning to the EEG data, we first examined the encoding of sample information in multivariate ERP patterns using RSA (see Materials and Methods). Specifically, we examined in each of the two

tasks (single-stream and dual-stream) the extent to which RSA patterns encoded (i) the concrete digit that was shown as sample stimulus (e.g., "4" or "8"), (ii) its color (i.e., red or blue), and (iii) the numerical magnitude information in a sample (i.e., 1-9, in terms of a numerical distance effect; see Materials and Methods and Figure 3.2*a*).

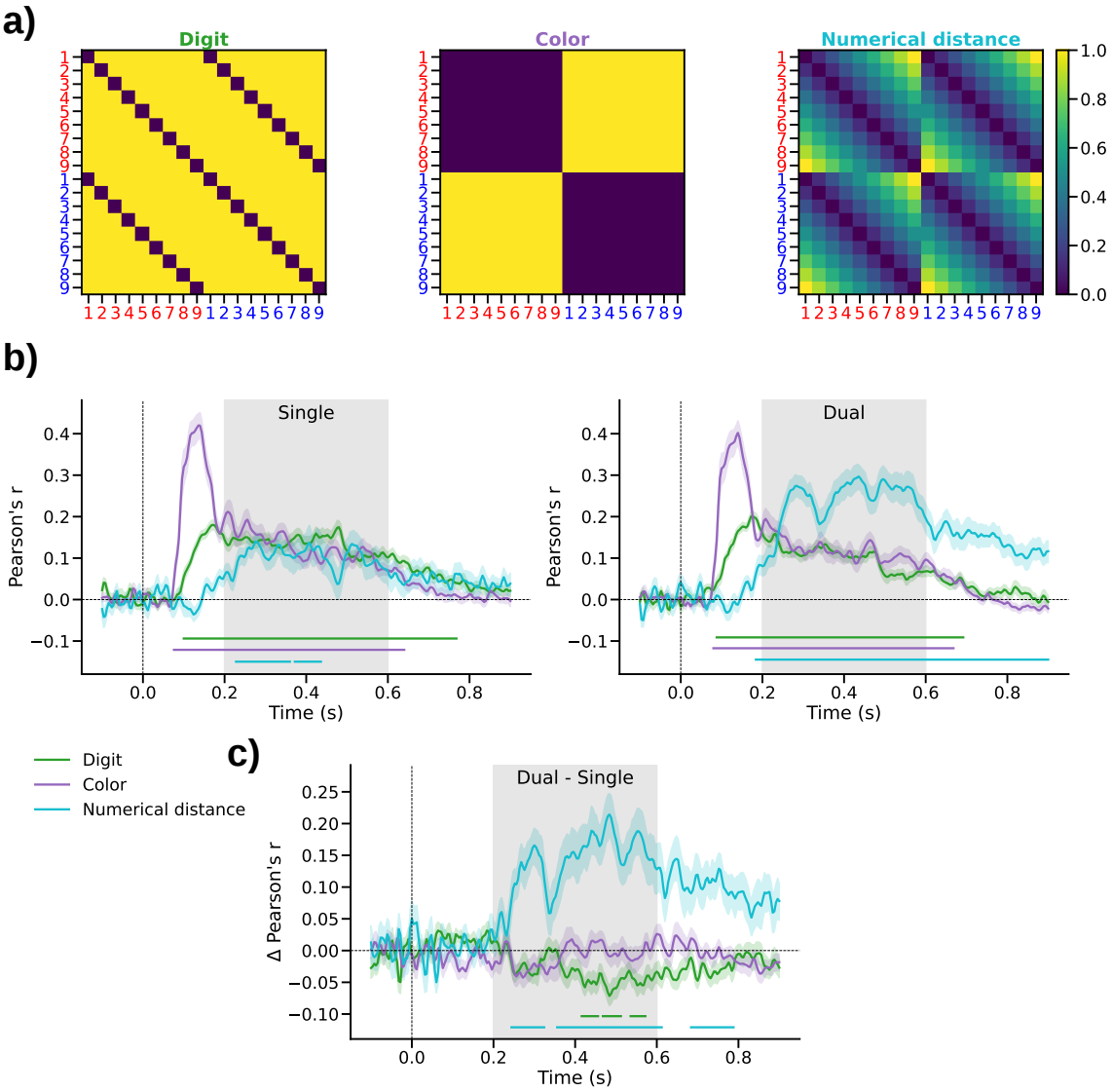

Figure 3.2. RSA results. a) Model RDMs encoding individual sample attributes. *Left*, "Digit" model encoding the unique number symbols; *Middle*, "Color" model encoding whether a sample was red or blue; *Right*, "Numerical distance" model encoding the samples' magnitude (1-9). For visual clarity, the model RDMs are illustrated before orthogonalization (see Materials and Methods). **b)** Correlation between orthogonalized model RDMs and the empirical ERP-RDMs in the single-stream (*left*) and dual-stream (*right*) task. Colored shadings show SE. Marker lines on bottom indicate time windows of significant differences from zero. Gray shading outlines the time window used in subsequent neurometric analysis (see Figure 3.4 below). **c)** Difference in correlation between the single- and dual-stream tasks. Same conventions as in *b*.

The visual attributes of a sample (i.e., its color and digit shape) were encoded early on, from approximately 100 to 700 ms after sample onset, in both tasks (Figure 3.2*b* all $p_{cluster}$<0.001). From approx. 200 ms on, the RSA patterns also encoded the samples' numerical magnitude, in terms of a significant numerical distance effect (single-stream: $p_{cluster}$<0.001; dual-stream: $p_{cluster}$<0.001), thus replicating and extending previous work (Appelhoff et al., 2022; Luyckx et al., 2019; Sheahan et al., 2021; Spitzer et al., 2017; Teichmann et al., 2018). Descriptively, the

numerical distance effect observed in the single-stream task, while robustly significant, appeared weaker than that in the dual-stream task.

Comparing the RSA time courses between the two task conditions (Figure 3.2*c*) confirmed that the numerical distance effect in the dual-stream task was significantly stronger ($p_{cluster}$<0.002). Surprisingly, we found no difference in color encoding between the two tasks, even though color was task-relevant only in the dual-stream task, but not in the single-stream task. Instead, in the single-stream task, we observed a relatively stronger representation of the concrete digit (i.e., the unique number symbol) that had been displayed ($p_{cluster}$<0.005). This effect was evident in a relatively late time window (approximately 400-600 ms, that is, only after the early visual encoding of digits and color).

Together, multivariate ERP patterns in both task conditions robustly encoded information about the sample's color, the number symbol it showed, and its numerical magnitude. The representation of samples in the single-stream task, however, showed qualitative differences, in terms of a relatively weaker encoding of numerical magnitude, and a relatively stronger encoding of the concrete sample stimuli.

Neurometric RSA results

Next, we examined potential distortions of the "neural numberline" underlying the neural magnitude representation disclosed in the above RSA results. To this end, we parameterized the numerical distance model (Figure 3.2*d*, *right*) to reflect distortions by k (compression/ anti-compression) and b (bias towards/against larger numbers), analogously as in our psychometric model (see Materials and Methods, Equation 3.1). We then used exhaustive gridsearch to determine for each participant the parameter combination with which the model fitted the data best. Figure 3.3*b* illustrates the improvement in fit (in terms of Δ r relative to the standard model with $k = 1$ and $b = 0$; cf. Figure 3.2*d*, *right*) in a representative time window (0.2-0.6 s; cf Figure 3.2*c* and Figure 3.3*a*). Note that neurometric mapping was performed using a log scale of k (where $log(k) = 0$ corresponds to $k = 1$, see Figure 3.3*b*) to avoid fitting bias (see Materials and Methods).

The results (Figure 3.3*b*) replicated our previous finding of a significant distortion of the neural numberline in the dual-stream task (Figure 3.3*b*, *right*). Specifically, like in our previous work (Appelhoff et al., 2022; Spitzer et al., 2017), we observed neurometric estimates of $k > 1$ (i.e., anti-compression) and of $b > 0$ (i.e., a bias towards larger numbers; p=0.013, FDR-corrected), which mirrors the pattern observed in the behavioral data (cf. Figure 3.1*c-d*, *orange*). However, contrary to our expectations, we found no evidence for a neurometric compression in the single-stream task (Figure 3.3*b*, *left*), where the psychometric weighting in behavior was clearly compressed (cf. Figure 3.1*c-d*, *blue*). Descriptively, the neurometric map in the single-stream task indicated a pattern similar to that in the dual-stream task (i.e., anti-compression $k > 1$, and positive bias, $b > 0$). However, the improvement in fit over the linear/unbiased model was only weak and statistically non-significant (Figure 3.3*b*, *left*) potentially reflecting that the numerical distance effect in the single-stream task was overall weaker (cf. Figure 3.2*b*).

Examining the mean neurometric parameter estimates in the two tasks statistically, they showed significant anti-compression ($k > 1$) both in the dual-stream [M=3.05, SE=0.40; t(29)=5.16, p<0.001, d=0.94] and in the single-stream task [M=3.81, SE=0.57; t(29)=4.90, p<0.001, d=0.89, t-tests against 1]. Direct comparison of neurometric k between the two tasks showed no significant difference [t(29)=1.25, p=0.22, d=0.28, paired t-test]. A positive offset bias (b) was evident in the dual-stream task [M=0.18, SE=0.04; t(29)=3.97, p<0.001, d=0.73] but not in the single-stream

task [M=0.04, SE=0.05; t(29)=0.75, p=0.46, d=0.14, t-tests against 0; difference between tasks t(29)=-2.27, p=0.03, d=0.54, paired t-test]. Together, the neurometric RSA results yielded no evidence for a compression of numerical magnitude akin to that observed in behavior in the single-stream task. If anything, the results were suggestive of anti-compression ($k > 1$) in both tasks, although it should be noted that the improvement in model fit (relative to a linear model) in the single-stream task was small and not statistically significant (Figure 3.3*b*, *left*).

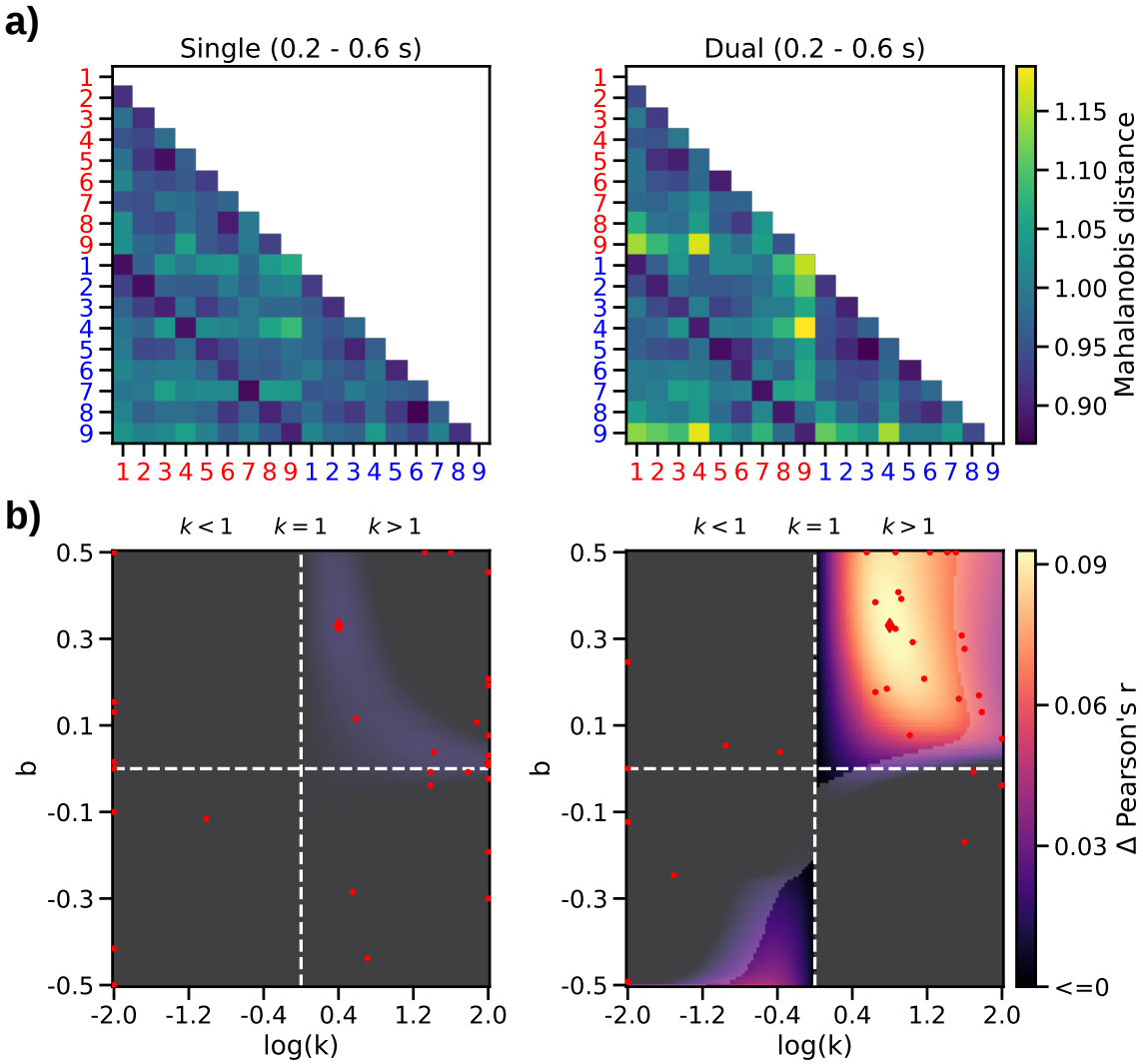

Figure 3.3. Neurometric RSA results. a) Mean ERP-RDMs in the time window of the numerical distance effect (see gray shading in Figure 3.2*b*). **b)** Mean neurometric maps. *Left*: single-stream task; *right*: dual-stream task. Dashed lines indicate linear ($k = 1$) and unbiased ($b = 0$) parameterizations. Color scale indicates increase in model correlation (Δ r) relative to the standard model where $k = 1$ and $b = 0$. Transparency mask delineates where the increase was statistically significant (p<0.05, FDR corrected). Red markers show maxima (diamond, mean map; dots, individual participant maps).

Univariate ERP results (CPP/P3)

We complemented our analysis by examining neurometric distortions also in univariate ERP signals, specifically in the sample-evoked CPP/P3 response. Previous research has implicated the CPP/P3 in decision formation, with its amplitude reflecting the perceived strength of evidence (Herding et al., 2019; O'Connell et al., 2012; Pisauro et al., 2017; Spitzer et al., 2016; Twomey et al., 2015; Wyart et al., 2015). CPP/P3 amplitudes were also found to be modulated by numerical sample values in the context of a comparison task (Spitzer et al., 2017).

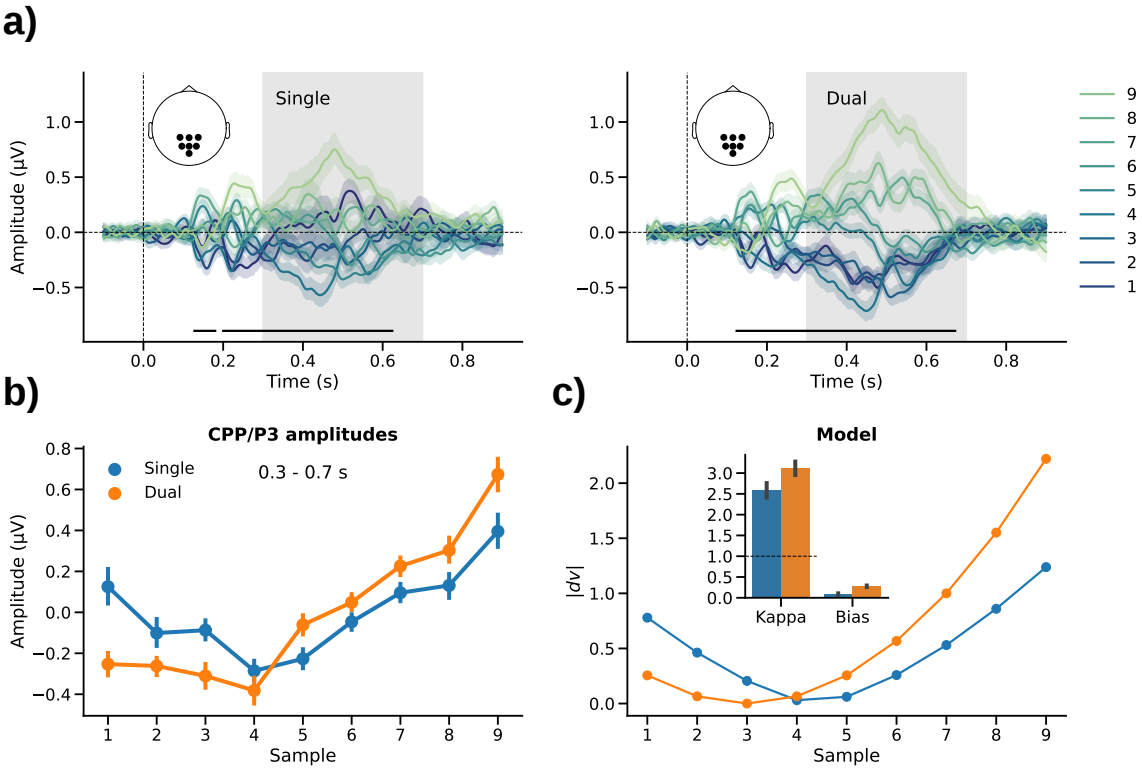

Figure 3.4. Univariate CPP/P3 results. a) Centro-parietal ERPs (mean-subtracted) evoked by sample values 1-9 in the single- (*left*) and dual-stream (*right*) tasks. Marker lines on bottom indicate time windows of significant differences between the 9 different sample values (all $p_{cluster}$<0.001, repeated measures analysis of variance). **b)** Mean CPP/P3 amplitudes averaged over the time window outlined by gray shading in a (0.3 - 0.7 s). **c)** Neurometric model fit of CPP/P3 amplitudes. Line plot shows grand mean fit. Inset bar graph shows mean parameter estimates. Error indicators in all panels show SE.

We observed modulations of CPP/P3 amplitude by numerical value both in the single-stream (Figure 3.4*a*, *left*; $p_{cluster}$<0.001) and in the dual-stream task (Figure 3.4*a*, *right*; $p_{cluster}$<0.001, repeated measures analyses of variance), with the modulation in the dual-stream task appearing descriptively stronger. The mean amplitudes showed a U-shaped pattern over numbers 1-9 (Figure 3.4*b*), consistent with previous findings that CPP/P3 reflects the strength of decisional evidence in an unsigned fashion (i.e., a theoretical quantity similar to the absolute $|dv|$ in our psychometric model, which would reflect the strength of evidence for either choice, "<" or ">"; see also Herding et al., 2019; O'Connell et al., 2012; Pisauro et al., 2017; Spitzer et al., 2016; Twomey et al., 2015; Wyart et al., 2015). Thus, for model-based analysis, we fitted Equation 3.1 to the CPP/P3 amplitude data (using pairwise distance matrices and gridsearch analogous to our neurometric RSA above), but using $|dv|$ to generate the model-predicted pattern.

The best-fitting parameter estimates are shown in Figure 3.4*c* (inset bar graph). Mirroring the RSA results, the parameter estimates based on CPP/P3 amplitude showed significant anti-compression ($k > 1$) in both tasks [single-stream: M=2.58, SE=0.55; t(29)=2.88, p=0.007, d=0.53, dual-stream: M=3.12, SE=0.51; t(29)=4.15, p<0.001, d=0.76, t-tests against 1], with no significant difference between tasks [t(29)=-0.69, p=0.49, d=0.18, paired t-test]. A positive offset bias was again observed in the dual-stream task [M=0.28, SE=0.06; t(29)=5.00, p<0.001, d=0.912], but not in the single-stream task [M=0.09, SE=0.05; t(29)=1.78, p=0.09, d=0.33, t-tests against 0; difference: t(29)=-2.97, p=0.006, d=0.63, paired t-test]. Together, the univariate ERP analysis thus corroborates our RSA finding that the neural processing of number samples in single-stream averaging was not characterized by compression but – if anything – by anti-compression, despite

the evidence for compression in subsequent behavioral choice (cf. Figure 3.1*c-d*).

3.3 Discussion

We observed opposite types of psychometric distortions (compression or anti-compression) in behavior when participants were instructed to process an identical stream of numbers in two distinct ways, namely, comparing the complete stream against a fixed target value or comparing two substreams against each other. However, contrary to expectations based on past research, the neural signals associated with the processing of the individual number samples showed evidence for anti-compression (i.e., enhanced processing of extreme values) under both instructions, regardless of whether extreme values were over- or underweighted in subsequent behavioral choice. We further observed qualitative differences between the sample representations in the two tasks, with a relatively weaker encoding of the samples' numerical magnitude in the single-stream task.

In psychophysical research concerned with the how attributes of a physical stimulus (e.g., size, weight, color) relate to their subjective experience or perception, human observers are commonly found to underweight extreme values. Such subjective "compression" of magnitude can be observed in a great variety of settings, from basic sensory-perceptual judgments (Fechner, 1860; Stevens, 1957) to economic decisions (Kellen et al., 2016; McAllister & Tarbert, 1999; Tversky & Kahneman, 1992). There exist various theoretical accounts for the origin and potential benefits of subjective compression in perception and decision making (Bhui & Gershman, 2018; Ciranka et al., 2022; de Gardelle & Summerfield, 2011; Li et al., 2017; Stewart et al., 2006; Summerfield & Li, 2018; Vandormael et al., 2017). Our present results seem to contrast these vast literatures with the finding that human brain signals tended to reflect the magnitude of numerical values in an anti-compressed fashion, even when they were compressed in later choice.

However, while the typical finding in many tasks is compression, there are task contexts where observers overweight extreme samples in choice behavior, in line with an anti-compression of sample values (Clarmann von Clarenau et al., 2022; Kunar et al., 2017; Ludvig et al., 2014; Luyckx et al., 2019; Shevlin et al., 2022; Spitzer et al., 2017; Tsetsos et al., 2012; Vanunu et al., 2020). We recently showed that such anti-compression can be beneficial in tasks that are computationally challenging (like our dual-stream comparison task), and where capacity-limited observers may be forced to selectively focus on a subset of the samples at the expense of others (Clarmann von Clarenau et al., 2022). Our present findings may suggest that in task contexts in which higher-level processing capacities are exceeded, participant behavior may more directly reflect the brain's default response to extreme values (i.e., privileged processing).

Despite the lack of evidence for different neural geometries of numerical magnitude in the two tasks, the samples' overall representation in neural signals yet differed. In the dual-stream task, neural signals encoded the numerical magnitude of a sample more strongly than in the single-stream task. This result was unexpected because nominally, the numerical magnitude of a sample was of equal relevance in both tasks. In the single-stream task, in turn, we found relatively stronger encoding of which unique number symbol was presented. Further, although the color of a sample (red/blue) was task-irrelevant in the single-stream task, the neural encoding of color was as strong and as sustained as in the color-based dual-stream task. While these results were unexpected, they may suggest more general differences in the role of "abstract" magnitude processing (Dehaene, 2003; Nieder, 2005; Nieder & Dehaene, 2009; Piazza & Izard, 2009; Walsh, 2003) in the two tasks. Potentially, in the more challenging dual-stream task participants relied more directly on an

intuitive "sense of magnitude" (Leibovich et al., 2017; Piazza et al., 2006; Spitzer et al., 2014) to gauge a sample's decision value. The computationally simpler single-stream task, in contrast, may have allowed them to engage in more symbolic-analytic processing (e.g., approximate arithmetics and/or verbalization) – processes that might be less amenable to EEG-decoding than the numerical distance pattern that prevailed in the dual-stream task.

The neural anti-compression of sample values in both of the tasks was also evident in univariate CPP/P3 signals which had previously been implicated in the decisional evaluation of stimulus information (Herding et al., 2019; O'Connell et al., 2012; Pisauro et al., 2017; Spitzer et al., 2016; Twomey et al., 2015; Wyart et al., 2015). Interpreting the amplitude of CPP/P3 signal as an index of the perceived strength of evidence, our findings in the single-stream task show a mismatch between the pattern observed in sample-by-sample processing (anti-compression, as revealed by neural data) and that in eventual judgment of the aggregate stream (compression, as evident in behavior). Future work will be required to identify the neural mechanisms underlying the eventual downweighting of extreme values in such task contexts, despite enhanced encoding in sample-level decision signals. It should be noted that the amplitude of P3 signals is known to be modulated by various factors, including how rare or surprising an event is (Donchin, 1981; DuncanJohnson & Donchin, 1977). However, the sample values in our experiment were uniformly distributed, that is, each value occurred equally often on average, ruling out an alternative explanation in terms of stimulus frequency.

Whether extreme values are over- or underweighted has major implications for the behavioral choices people make. One and the same numerical evidence, as in the present study, may lead to opposite choices, depending on how people respond to and process extreme values. For instance, anti-compression can explain systematic violations of "rational" axioms, such as transitivity, in multi-attribute choice (Summerfield & Tsetsos, 2015; Tsetsos et al., 2016). More generally, non-linear distortions of objective data (such as numbers) have often been interpreted as paradigmatic manifestations of seemingly "irrational" human behaviors in decision making (e.g., non-linear probability weighting in risky choice as assumed in cumulative prospect theory; Tversky & Kahneman, 1992). However, over the past years a new literature has evolved that recasts these behaviors as well-adapted policies of capacity-limited observers, fostering rather than hampering their measurable performance under these constraints (Bhui et al., 2021; Gigerenzer & Brighton, 2009; Gigerenzer et al., 2011; Juechems et al., 2021; Lieder & Griffiths, 2020; Sims, 2003, 2010; Tsetsos et al., 2016). Here, we shed new light on the open question of how such adaptive distortions may arise mechanistically, in terms of the neural signal patterns evoked by samples of evidence while observers are in the process of reaching a decision. Our finding of a "default" anti-compression of values in neural responses – regardless of subsequent behavior – raises the question at which exact processing stage value compression emerges, and how it leads to, for instance, the well-known "diminishing sensitivity" to larger values in economic choices (e.g., Tversky & Kahneman, 1992).

3.4 Materials and Methods

3.4.1 Participants

Thirty-two healthy volunteers took part in the experiment. We excluded two participants who reported having misunderstood the task instructions and who performed near chance level (50% correct choices) in one of the tasks (52% and 53%, respectively; both $p>0.4$, Binomial tests against 0.5). Results are reported for the n=30 remaining participants (15 male, 15 female; mean age 27.4

± 4.9 years; one left handed). All participants provided written informed consent and received €10 per hour as compensation, in addition to a €10 flat fee for participation, as well as a performance-dependent bonus (€7.03 ± 1.08 on average). The study was approved by the ethics committee of the Max Planck Institute for Human Development.

3.4.2 Experimental design

Each participant performed two variants of a sequential number integration task (Clarmann von Clarenau et al., 2022; Spitzer et al., 2017). The stimulus protocols in the two task variants were identical (Figure 3.1*a*). On each trial, participants viewed a sequence of 10 Arabic digits (randomly drawn from a uniform distribution of numbers 1 to 9) displayed in either red or blue font color (randomly assigned to each sample, with the restriction that each sequence contained 5 red and 5 blue samples). In the "averaging" (single-stream) task, participants were asked to judge whether the average of all 10 number samples in the sequence (regardless of their color) was larger or smaller than 5. In the "comparison" (dual-stream) task, participants were asked to indicate whether the red or the blue samples had the higher average value. Past behavioral work has shown a psychometric compression of number values in the single-stream task, whereas anti-compression was evident in the latter dual-stream task (Clarmann von Clarenau et al., 2022; Spitzer et al., 2017).

The experiment was programmed in Python using the PsychoPy package (Peirce et al., 2019) and run on a Windows 10 PC. The experiment code is available on GitHub (https://github.com/sappelhoff/ecomp_experiment). Throughout the experiment, we additionally recorded eye-movements using an EyeLink 1000 Plus (SR Research Ltd., Canada), which were not analyzed in the present study. Participants were informed about the eye-movement recording and were instructed to keep their gaze at the center of the screen throughout the experiment.

Each trial started with a white central fixation stimulus (a combination of bulls eye and cross hair; Thaler et al., 2013) on an otherwise black screen. After 500 ms, the fixation stimulus disappeared and the number sequence was presented at a rate of 350 ms per sample (font Liberation Sans; height 3° visual angle; see Figure 3.1*a*). Each sample was smoothly faded to black after 270 ms to improve the visual experience of the stimulus transitions. After the last sample, participants were prompted to enter a response by pressing the left or right button on a USB response pad (The Black Box ToolKit Ltd., UK). To avoid left/right motor response preparation during sequence presentation, in each of the two tasks, we randomized the mapping of responses ("smaller" or "larger" in the averaging task; "red" or "blue" in the comparison task) onto left/right button presses trial-by-trial, using a response screen (Figure 3.1*a*, *right*).

If participants failed to respond within 3 s, the trial was discarded and after a delay of 100 ms, a message ("too slow!") was displayed in red color for 1 s. On average, participants responded within 0.67±0.27 s and timeouts occurred only on 0.03 % of trials. On the remaining trials, performance feedback was displayed ("correct" or "wrong", in green or orange color, respectively) for 350 ms. All feedback was displayed centrally in Liberation Sans font with a height of 1° visual angle. On 4.85% of trials, in which the objective sequence average was precisely 5 (in the single-stream task) or identical for red and green samples (in the dual-stream task), a random feedback message was displayed. These trials were excluded from the analysis of accuracy levels, but were included in the modeling- and EEG analyses (see also Spitzer et al., 2017). After feedback, the central fixation stimulus re-appeared and after 500 to 1500 ms (randomly varied), the next trial started.

Each participant first performed 300 trials in one of the tasks (single-stream averaging or dual-stream comparison), followed by 300 trials in the other task (in counterbalanced serial order across

subjects). Thus, 3000 number samples were presented in each task condition and participant. Trials were performed in blocks of 50, with summary performance feedback (percentage correct choices) being provided after each block. After completing all blocks of the first task, participants received the instructions for the second task. To avoid differences in stimulus input, the second task was performed on the exact same number sequences as the first task. Upon completing the second task, participants received a monetary bonus depending on their mean accuracy in both tasks.

3.4.3 EEG recording

The experiment was performed in an electrically shielded and soundproof cabin. Scalp EEG was recorded with 64 active electrodes (actiCap, Brain Products GmbH Munich, Germany) positioned according to the international 10% system. Electrode FCz was used as the recording reference. We additionally recorded the horizontal and vertical electrooculogram (EOG) and electrocardiogram (ECG) using passive electrode pairs with bipolar referencing. All electrodes were prepared to have an impedance of less than 10 kΩ. The data were recorded using a BrainAmp DC amplifier (Brain Products GmbH Munich, Germany) at a sampling rate of 1000 Hz, with an RC high-pass filter with a half-amplitude cutoff at 0.016 Hz (roll-off: 6 dB/octave) and low-pass filtered with an anti-aliasing filter of half-amplitude cutoff 450 Hz (roll-off: 24 dB/octave). The full dataset will be made available upon publication of this article on GIN: https://gin.g-node.org/sappelhoff/mpib_ecomp_sourcedata/.

3.4.4 Behavioral data analysis

We calculated model-free decision weights to examine how strongly each numerical sample value (1, 2, ..., 9) contributed to participants' choices in the two tasks. In the single-stream task, these weights were computed as the proportion of times a sample value was associated with the subsequent choice "larger". Analogously, in the dual-stream task, the weights were computed as the proportion of times the sample's color (i.e., red or blue) was subsequently chosen (see also Spitzer et al., 2017). For comparison with model predictions (see below), we computed decision weights also from the model-predicted choice probabilities (CP, see Equation 3.3) obtained from using the best fitting parameter estimates in each participant.

3.4.5 Psychometric model

To quantify psychometric distortions (i.e., compression or anti-compression) in behavior, we used a simple psychometric model that has been used extensively in previous work (Appelhoff et al., 2022; Clarmann von Clarenau et al., 2022; Li et al., 2017; Luyckx et al., 2019; Spitzer et al., 2017). The model formalizes the transformation of objective sample values X (here: numbers 1-9, normalized to the range [-1, 1]) into a subjective decision value dv as a sign-preserving power function:

$$dv = \frac{X + b}{|X + b|} \times |X + b|^k, \tag{3.1}$$

where exponent k (kappa) determines the overall shape of the transformation ($k < 1$: compression; $k = 1$: linear; $k > 1$: anti-compression). Parameter b (bias) implements an overall weighting bias towards smaller ($b < 0$) or larger ($b > 0$) numbers. Sample-level decision values (dv) are integrated into a trial-level decision value (DV) by summation over samples:

$$DV = \sum_{i=1}^{10} dv_i \times c_i, \tag{3.2}$$

where c is an indicator variable denoting a sample's color (red: -1, blue: +1) in the dual-stream task. In the single-stream task, c was fixed at 1. This way, Equation 3.2 effectively implements a comparison between streams in the dual-stream task, and simple averaging in the single-stream task. Finally, the trial level decision value (DV) is transformed into a choice probability according to a logistic function:

$$CP = \frac{1}{1 + e^{\frac{-DV}{s}}}, \tag{3.3}$$

where CP denotes the probability of choosing ">5" (in the single-steam task) or "blue>red" (in the dual-stream task), and parameter s quantifies the level of decision noise, with larger values of s implying more random choices.

The model was fitted to each participant's individual choice data using the Nelder-Mead method as implemented in SciPy (Virtanen et al., 2020), with parameter values restricted to the ranges (k: [0, 5]; b: [-0.5, 0.5], s: [0.01, 3]). Fitting was performed iteratively using 900 combinations of different starting values for each task condition, and the solution with the lowest Bayesian Information Criterion (BIC) was used in the analysis. Statistical analysis of the fitted parameters proceeded with conventional inferential tests on the group level.

3.4.6 EEG preprocessing

We used functions from MNE-Python (Gramfort et al., 2013) and PyPrep (based on Bigdely-Shamlo et al., 2015) to automatically mark noisy segments and bad channels in the EEG recordings. We additionally screened all recordings visually to reject noisy segments or bad channels that the automatic procedures had missed. This way, on average, 2.5 ± 1.6 channels were discarded per participant. Next, we corrected ocular and cardiac artifacts using independent component analysis (ICA). To this end, we high-pass filtered a copy of the raw data at 1 Hz and downsampled it to 100 Hz. We then ran an extended infomax ICA on all EEG channels and time points that were not marked as bad in the prior inspection. Using the EOG and ECG recordings, we identified stereotypical eye blink, eye movement, and heartbeat artifact components through correlation with the independent component time courses. We visually inspected and rejected the artifact components before applying the ICA solution to the (Winkler et al., 2015). We then filtered the ICA-cleaned data between 0.1 and 40 Hz, interpolated bad channels, and re-referenced each channel to the average of all channels.

3.4.7 Event-related potentials (ERPs)

We epoched the preprocessed data from -0.1 to 0.9 s relative to each sample stimulus onset. Remaining bad epochs were rejected using a thresholding approach from the FASTER pipeline (Step 2; Nolan et al., 2010). On average, n=5764 clean epochs (96.1%) per participant were retained for analysis. The epochs were then downsampled to 250 Hz and baseline corrected relative to the period from -0.1 to 0 s before stimulus onset. Since our analyses focused on stimulus-specific effects, we subtracted the overall mean waveform from the individual epochs, in each of the two task conditions. The mean-subtracted epochs were then averaged into stimulus-specific ERPs for

each sample value (1, 2, ..., 9) in each color (red/blue). Note that the individual samples in a stream were statistically independent by design, allowing us to examine stimulus-specific ERP responses in a time window that overlapped with the onset of the next sample stimulus.

3.4.8 Representational similarity analysis

We used representational similarity analysis (RSA) to examine the encoding of sample information in multivariate ERP patterns (Kriegeskorte & Kievit, 2013). Specifically, we examined the representational geometry of our stimulus space (numbers 1 to 9, colored red or blue) in terms of the multivariate (dis-)similarity between the ERP topographies (64 channels) associated with the 18 different stimuli. Representational dissimilarity was computed as the Mahalanobis distance, between each pair of stimuli, yielding an 18x18 representational dissimilarity matarix (RDM), at each time point of the peri-sample epoch. To compute the Mahalanobis distance, we fitted a general linear model to the z-scored trial data, with each stimulus type specified as a condition, and used the residual trial-by-trial variance for pairwise distance calculation. This procedure ensured multivariate noise normalization for the RDMs (Guggenmos et al., 2018). Below, we refer to the thus obtained RDMs as ERP-RDMs.

To examine the information encoded in the ERP-RDM time courses, we used three different model RDMs (see Figure 3.2*a*) reflecting (i) the unique digit symbols, with minimum dissimilarity between identical digits, and maximum dissimilarity between distinct digits ("digit" model), (ii) the samples' color, with minimum (maximum) dissimilarity between same (different) colors ("color" model), and (iii) the numerical distance between samples, that is, the arithmetic difference between their objective number values ("numerical distance" model). To render the three models fully independent, we recursively orthogonalized each model RDM with respect to all others using the Gram-Schmidt process (Appelhoff et al., 2022; Spitzer et al., 2017). Finally, we assessed the match between each model RDM and the empirically observed ERP-RDMs via Pearson correlation at each time point, using only the lower triangle of the RDMs and omitting the diagonal, to exclude redundant matrix cells.

For statistical analysis of the RSAs time courses, we used t-tests against zero with cluster-based permutation testing to control for multiple comparisons over time points (Maris & Oostenveld, 2007). To test whether RSA results differed between the single- and dual-stream tasks, we first computed their difference, followed by cluster-based permutation tests against zero. All permutation tests were performed over 1000 iterations with a cluster-defining threshold of p=0.01 and cluster length as the critical statistic (thresholded at p=0.01).

3.4.9 Analysis of neurometric distortions

The theoretical model underlying conventional RSA of numerical distance effects (see above) is a straight number line, where the numbers (1, 2, ..., 9) are equidistant. The standard numerical distance model (Figure 3.2*a*, *right*) is equivalent to a model of *dv* according to Equation 3.1 (see Psychometric model) where $k = 1$ and $b = 0$. To examine potential nonlinear distortions of the number representations in neural signals ("neurometric" distortions), we constructed numerical distance models based on *dv* while varying the values of k (from $log(k) = -2$ to $+2$, see Figure 3.3*b*) and b (from -0.5 to 0.5). Varying k on a log scale centered around $log(k) = 0$ (i.e., $k = 1$) ensured that parameter estimates were not biased to show anti-compression (or compression) by chance in subsequent gridsearch. For each parameter combination, we correlated the resulting model RDMs with the ERP-RDM, yielding a grid ("neurometric map") of the parameter space (see

Figure 3.3*b*). The parameter combination with the maximum correlation was used as the estimate of the participant's neurometric distortion parameters (see also Appelhoff et al., 2022; Spitzer et al., 2017). Statistical analysis of the neurometric parameter estimates proceeded with conventional statistical tests on the group level.

3.4.10 Univariate ERP analysis

For complementary inspection of univariate ERP responses evoked by the number samples, we examined the stimulus-specific ERP (see above) for each sample value (1-9; collapsed across red/blue colors). To focus on CPP/P3 responses (see Results), the ERPs were pooled over centro-parietal channels (CP1, P1, POz, Pz, CPz, CP2, P2) and amplitudes were examined in a time window from 300 ms to 700 ms based on previous work (Appelhoff et al., 2022; Polich, 2007; Spitzer et al., 2017; Wyart et al., 2015). The ERP time courses were analyzed statistically using cluster-based permutation testing (see above). For model-based analysis, we used the same approach as in our analyses of neurometric distortions in RSA (see above), except that the model RDMs were constructed from $|dv|$ (i.e., the absolute, unsigned magnitude of dv, see Results) and correlated with the pairwise differences in univariate ERP amplitude between samples 1-9.

3.5 Supplemental material

All data will be made available upon publication of this article on GIN: https://gin.g-node.org/sappelhoff/mpib_ecomp_sourcedata/.

All analysis code will be made available upon publication of this article on GitHub: https://github.com/sappelhoff/ecomp_analysis. The experiment code is available on GitHub: https://github.com/sappelhoff/ecomp_experiment.

3.6 Ethics information

The study was approved by the ethics committee of the Max Planck Institute for Human Development.

3.7 Acknowledgments

We thank Anna Faschinger, Gabriele Inciuraite, Aleksandra Zinoveva, Larissa Samaan, Simon Ciranka, and Jann Wäscher for help with data collection, and Verena Clarmann von Clarenau and Thorsten Pachur for helpful discussions.

3.8 Author contributions

SA, BS: Formal analysis, Visualization, Methodology, Conceptualization, Project Administration, Writing - original draft

SA: Investigation, Validation, Data curation, Software

SA, BS, RH: Writing - review & editing

BS: Supervision

RH, BS: Resources

Corresponding author: Stefan Appelhoff

For a more detailed overview of author contributions, please see Appendix B.

3.9 Funding

This work was supported by a European Research Council Consolidator Grant ERC-2020-COG-101000972 (BS).

3.10 Competing interests

The authors declare no competing interests.

3.11 References

Appelhoff, S., Hertwig, R., & Spitzer, B. (2022). Control over sampling boosts numerical evidence processing in human decisions from experience. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhac062

Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica*, *22*(1), 23. https://doi.org/10.2307/1909829

Bhui, R., & Gershman, S. J. (2018). Decision by sampling implements efficient coding of psychoeconomic functions. *Psychological Review*, *125*(6), 985–1001. https://doi.org/10.1037/rev0000123

Bhui, R., Lai, L., & Gershman, S. J. (2021). Resource-rational decision making. *Current Opinion in Behavioral Sciences*, *41*, 15–21. https://doi.org/10.1016/j.cobeha.2021.02.015

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., & Robbins, K. A. (2015). The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, *9*. https://doi.org/10.3389/fninf.2015.00016

Ciranka, S., Linde-Domingo, J., Padezhki, I., Wicharz, C., Wu, C. M., & Spitzer, B. (2022). Asymmetric reinforcement learning facilitates human inference of transitive relations [Publisher: Nature Publishing Group]. *Nature Human Behaviour*, 1–10. https://doi.org/10.1038/s41562-021-01263-w

Clarmann von Clarenau, V., Pachur, T., & Spitzer, B. (2022). *Over- and Underweighting of Extreme Values in Decisions from Sequential Samples* (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/6yj4r

de Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences*, *108*(32), 13341–13346. https://doi.org/10.1073/pnas.1104517108

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences*, *7*(4), 145–147. https://doi.org/10.1016/s1364-6613(03)00055-x

Donchin, E. (1981). Surprise!? Surprise? *Psychophysiology*, *18*(5), 493–513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x

DuncanJohnson, C. C., & Donchin, E. (1977). On Quantifying Surprise: The Variation of Event-Related Potentials With Subjective Probability. *Psychophysiology*, *14*(5), 456–467. https://doi.org/10.1111/j.1469-8986.1977.tb01312.x

Fechner, G. (1860). *Elemente der Psychophysik*. Breitkopf und Härtel. https://books.google.de/books?id=6rINAAAAYAAJ

Gigerenzer, G., & Brighton, H. (2009). Homo Heuristicus: Why Biased Minds Make Better Inferences [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1756-8765.2008.01006.x]. *Topics in Cognitive Science*, *1*(1), 107–143. https://doi.org/10.1111/j.1756-8765.2008.01006.x

Gigerenzer, G., Hertwig, R., & Pachur, T. (2011). *Heuristics*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199744282.001.0001

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python [Publisher: Frontiers]. *Frontiers in Neuroscience*, *7*(267). https://doi.org/10.3389/fnins.2013.00267

Guggenmos, M., Sterzer, P., & Cichy, R. M. (2018). Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. *NeuroImage*, *173*, 434–447. https://doi.org/10.1016/j.neuroimage.2018.02.044

Herding, J., Ludwig, S., von Lautz, A., Spitzer, B., & Blankenburg, F. (2019). Centro-parietal EEG potentials index subjective evidence and confidence during perceptual decision making. *NeuroImage*, *201*, 116011. https://doi.org/10.1016/j.neuroimage.2019.116011

Juechems, K., Balaguer, J., Spitzer, B., & Summerfield, C. (2021). Optimal utility and probability functions for agents with finite computational precision [Publisher: National Academy of Sciences Section: Social Sciences]. *Proceedings of the National Academy of Sciences*, *118*(2). https://doi.org/10.1073/pnas.2002232118

Kellen, D., Pachur, T., & Hertwig, R. (2016). How (in)variant are subjective representations of described and experienced risk and rewards? *Cognition*, *157*, 126–138. https://doi.org/10.1016/j.cognition.2016.08.020

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412. https://doi.org/10.1016/j.tics.2013.06.007

Kunar, M. A., Watson, D. G., Tsetsos, K., & Chater, N. (2017). The influence of attention on value integration. *Attention, Perception, & Psychophysics*, *79*(6), 1615–1627. https://doi.org/10.3758/s13414-017-1340-7

Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2017). From sense of number to sense of magnitude: The role of continuous magnitudes in numerical cognition [Publisher: Cambridge University Press]. *Behavioral and Brain Sciences*, *40*. https://doi.org/10.1017/S0140525X16000960

Li, V., Castañón, S. H., Solomon, J. A., Vandormael, H., & Summerfield, C. (2017). Robust averaging protects decisions from noise in neural computations [Publisher: Public Library of Science]. *PLOS Computational Biology*, *13*(8), e1005723. https://doi.org/10.1371/journal.pcbi.1005723

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources [Publisher: Cambridge University Press]. *Behavioral and Brain Sciences*, *43*. https://doi.org/10.1017/S0140525X1900061X

Longo, M. R., & Lourenco, S. F. (2007). Spatial attention and the mental number line: Evidence for characteristic biases and compression. *Neuropsychologia*, *45*(7), 1400–1407. https://doi.org/10.1016/j.neuropsychologia.2006.11.002

Ludvig, E. A., Madan, C. R., & Spetch, M. L. (2014). Extreme Outcomes Sway Risky Decisions from Experience: Risky Decisions and Extreme Outcomes. *Journal of Behavioral Decision Making*, *27*(2), 146–156. https://doi.org/10.1002/bdm.1792

Luyckx, F., Nili, H., Spitzer, B., & Summerfield, C. (2019). Neural structure mapping in human probabilistic reward learning (D. Lee, J. I. Gold, D. Lee, & M. Chafee, Eds.) [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *8*, e42816. https://doi.org/10.7554/eLife.42816

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

McAllister, P., & Tarbert, H. (1999). Bargaining, utility and rents: Analysing the effect of potential lease termination on rent negotiation outcomes. *Journal of Property Investment & Finance*, *17*(4), 353–364. https://doi.org/10.1108/14635789910271746

Nieder, A. (2005). Counting on neurons: The neurobiology of numerical competence [Number: 3 Publisher: Nature Publishing Group]. *Nature Reviews Neuroscience*, *6*(3), 177–190. https://doi.org/10.1038/nrn1626

Nieder, A., & Dehaene, S. (2009). Representation of Number in the Brain. *Annual Review of Neuroscience*, *32*(1), 185–208. https://doi.org/10.1146/annurev.neuro.051508.135550

Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, *37*(1), 149–157. https://doi.org/10.1016/s0896-6273(02)01144-3

Nolan, H., Whelan, R., & Reilly, R. B. (2010). FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection. *Journal of Neuroscience Methods*, *192*(1), 152–162. https://doi.org/10.1016/j.jneumeth.2010.07.015

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans [Number: 12 Publisher: Nature Publishing Group]. *Nature Neuroscience*, *15*(12), 1729–1735. https://doi.org/10.1038/nn.3248

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Piazza, M., & Izard, V. (2009). How Humans Count: Numerosity and the Parietal Cortex. *The Neuroscientist*, *15*(3), 261–273. https://doi.org/10.1177/1073858409333073

Piazza, M., Mechelli, A., Price, C. J., & Butterworth, B. (2006). Exact and approximate judgements of visual and auditory numerosity: An fMRI study. *Brain Research*, *1106*(1), 177–188. https://doi.org/10.1016/j.brainres.2006.05.104

Pisauro, M. A., Fouragnan, E., Retzler, C., & Philiastides, M. G. (2017). Neural correlates of evidence accumulation during value-based decisions revealed via simultaneous EEG-fMRI. *Nature Communications*, *8*(1), 15808. https://doi.org/10.1038/ncomms15808

Polich, J. (2007). Updating P300: An Integrative Theory of P3a and P3b. *Clinical neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Sheahan, H., Luyckx, F., Nelli, S., Teupe, C., & Summerfield, C. (2021). Neural state space alignment for magnitude generalization in humans and recurrent networks. *Neuron*, *109*, 1214–1226. https://doi.org/10.1016/j.neuron.2021.02.004

Shevlin, B. R. K., Smith, S. M., Hausfeld, J., & Krajbich, I. (2022). High-value decisions are fast and accurate, inconsistent with diminishing value sensitivity. *Proceedings of the National Academy of Sciences*, *119*(6), e2101508119. https://doi.org/10.1073/pnas.2101508119

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, *50*(3), 665–690. https://doi.org/10.1016/S0304-3932(03)00029-1

Sims, C. A. (2010). Chapter 4 - Rational Inattention and Monetary Economics. In B. M. Friedman & M. Woodford (Eds.), *Handbook of Monetary Economics* (pp. 155–181). Elsevier. https://doi.org/10.1016/B978-0-444-53238-1.00004-1

Spitzer, B., Fleck, S., & Blankenburg, F. (2014). Parametric Alpha- and Beta-Band Signatures of Supramodal Numerosity Information in Human Working Memory. *Journal of Neuroscience*, *34*(12), 4293–4302. https://doi.org/10.1523/JNEUROSCI.4580-13.2014

Spitzer, B., Blankenburg, F., & Summerfield, C. (2016). Rhythmic gain control during supramodal integration of approximate number. *NeuroImage*, *129*, 470–479. https://doi.org/10.1016/j.neuroimage.2015.12.024

Spitzer, B., & Haegens, S. (2017). Beyond the Status Quo: A Role for Beta Oscillations in Endogenous Content (Re)Activation. *eNeuro*, *4*(4). https://doi.org/10.1523/ENEURO.0170-17.2017

Spitzer, B., Waschke, L., & Summerfield, C. (2017). Selective overweighting of larger magnitudes during noisy numerical comparison. *Nature Human Behaviour*, *1*(8), 1–8. https://doi.org/10.1038/s41562-017-0145

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*(3), 153–181. https://doi.org/10.1037/h0046162

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology*, *53*(1), 1–26. https://doi.org/10.1016/j.cogpsych.2005.10.003

Summerfield, C., & Li, V. (2018). Perceptual suboptimality: Bug or feature? [Publisher: Cambridge University Press]. *Behavioral and Brain Sciences*, *41*. https://doi.org/10.1017/S0140525X18001437

Summerfield, C., & Parpart, P. (2022). Normative Principles for Decision-Making in Natural Environments. *Annual Review of Psychology*, *73*(1), 53–77. https://doi.org/10.1146/annurev-psych-020821-104057

Summerfield, C., & Tsetsos, K. (2015). Do humans make good decisions? *Trends in cognitive sciences*, *19*(1), 27–34. https://doi.org/10.1016/j.tics.2014.11.005

Teichmann, L., Grootswagers, T., Carlson, T., & Rich, A. N. (2018). Decoding Digits and Dice with Magnetoencephalography: Evidence for a Shared Representation of Magnitude [Publisher: MIT Press]. *Journal of Cognitive Neuroscience*, *30*(7), 999–1010. https://doi.org/10.1162/jocn_a_01257

Thaler, L., Schütz, A. C., Goodale, M. A., & Gegenfurtner, K. R. (2013). What is the best fixation target? The effect of target shape on stability of fixational eye movements. *Vision Research*, *76*, 31–42. https://doi.org/10.1016/j.visres.2012.10.012

Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal [ISBN: 9781119569107 Publisher: National Academy of Sciences Section: Biological Sciences]. *Proceedings of the National Academy of Sciences*, *109*(24), 9659–9664. https://doi.org/10.1073/pnas.1119569109

Tsetsos, K., Moran, R., Moreland, J., Chater, N., Usher, M., & Summerfield, C. (2016). Economic irrationality is optimal during noisy decision making. *Proceedings of the National Academy of Sciences*, *113*(11), 3102–3107. https://doi.org/10.1073/pnas.1519157113

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/10.1007/BF00122574

Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience*, *42*(1), 1636–1643. https://doi.org/10.1111/ejn.12936

Vandormael, H., Herce Castañón, S., Balaguer, J., Li, V., & Summerfield, C. (2017). Robust sampling of decision information during perceptual choice. *Proceedings of the National Academy of Sciences*, *114*(10), 2771–2776. https://doi.org/10.1073/pnas.1613950114

Vanunu, Y., Hotaling, J. M., & Newell, B. R. (2020). Elucidating the differential impact of extreme-outcomes in perceptual and preferential choice. *Cognitive Psychology*, *119*, 101274. https://doi.org/10.1016/j.cogpsych.2020.101274

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity [Publisher: Elsevier]. *Trends in Cognitive Sciences*, *7*(11), 483–488. https://doi.org/10.1016/j.tics.2003.09.002

Winkler, I., Debener, S., Müller, K., & Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP [ISSN: 1558-4615]. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 4101–4105. https://doi.org/10.1109/EMBC.2015.7319296

Wyart, V., de Gardelle, V., Scholl, J., & Summerfield, C. (2012). Rhythmic Fluctuations in Evidence Accumulation during Decision Making in the Human Brain. *Neuron*, *76*(4), 847–858. https://doi.org/10.1016/j.neuron.2012.09.015

Wyart, V., Myers, N. E., & Summerfield, C. (2015). Neural Mechanisms of Human Perceptual Choice Under Focused and Divided Attention [Publisher: Society for Neuroscience Section: Articles]. *Journal of Neuroscience*, *35*(8), 3485–3498. https://doi.org/10.1523/JNEUROSCI.3276-14.2015

4 | In COM we trust: Feasibility of USB-based event marking

Stefan Appelhoff & Tristan Stenner

Appelhoff, **S.**, & Stenner, T. (2021). In COM we trust: Feasibility of USB-based event marking. *Behavior Research Methods*, *53*(6), 2450–2455. https://doi.org/10.3758/s13428-021-01571-z

Abstract

Modern experimental research often relies on the synchronization of different events prior to data analysis. One way of achieving synchronization involves marking distinct events with electrical pulses (event markers or "TTL pulses"), which are continuously recorded with research hardware, and can later be temporally aligned. Traditionally, this event marking was often performed using the parallel port in standard personal computers. However, the parallel port is disappearing from the landscape of computer hardware, being replaced by a serial (COM) port, namely the USB port. To find an adequate replacement for the parallel port, we evaluated four Microcontroller Units (MCUs) and the LabJack U3, an often-used USB data acquisition device, in terms of their latency and jitter for sending event markers in a simulated experiment on both Windows and Linux. Our results show that all four MCUs were comparable to the parallel port in terms of both latency and jitter, and consistently achieved latencies under 1 ms. With some caveats, the LabJack U3 can also achieve comparable latencies. In addition to the collected data, we share extensive documentation on how to build and use MCUs for event marking, including code examples. MCUs are a cost-effective, flexible, and performant replacement for the disappearing parallel port, enabling event marking and synchronization of data streams.

4.1 Introduction

Experimental research in a multitude of scientific disciplines involves the presentation of stimuli to research subjects. Be it displaying a visual stimulus to a human, applying an electric shock to a rodent, or switching off the lights in a room being navigated by an autonomous robot – researchers need to record these events together with other data to make sense of the results. A traditional implementation for event marking is the parallel port available on standard computers (Clerc et al., 2016). The parallel port works with direct current signals that directly translate to digital states on the receiving interface, and can thus achieve microsecond resolution of data transmission (Stewart, 2006; Voss et al., 2007). These fast direct current signals, also called Transistor-transistor logic (TTL) pulses, persuaded manufacturers of recording hardware for experimental research to adopt the parallel port as a standard interface for implementing event marking. However, the parallel port's original role as a general interface to transmit data has been taken over by serial communication protocols, specifically the ubiquitous Universal Serial Bus (USB) protocol. As parallel port interfaces are replaced by USB ports on commercially available computers, it is becoming increasingly difficult to obtain modern computer hardware that supports a parallel port "out of the box". Yet, most data recording hardware for experimental research still relies on event marker inputs sent from a parallel port interface. Consequently, researchers often find themselves hoarding outdated computer equipment or relying on workarounds such as PCI adapter cards or old docking stations for modern laptops. The problem of parallel port availability is even more pronounced for users of Apple computers, which are traditionally produced without a parallel port interface (Knight, 1997). But even for users of recent versions of Microsoft Windows, access to the parallel port has become increasingly difficult for native applications and impossible for web browser-based studies (see e.g., Bridges et al., 2020).

Workarounds for event marking should be treated with caution: because they tend not to be exhaustively tested, they can introduce dangerous uncertainty about the true latency between an event and its associated marker in the data. Some analysis techniques such as event-related potentials (ERPs, Luck, 2014) can tolerate a small *constant* latency of a few milliseconds, but quickly become unusable with unpredictable variations of the latencies (i.e., jitter). Commercial manufacturers are beginning to acknowledge this situation and to release products tailored to bridging the link from USB connectors to the proprietary connectors expecting TTL pulse inputs as from a parallel port. Currently, there are at least seven products to choose from (e.g., Canto et al., 2011, see Supplemental Material for a list), most of which have been tested by their respective manufacturers. Unfortunately, however, most of the products have at least one of the following drawbacks: (i) they are expensive, (ii) they are specific to a particular type of hardware, or (iii) they provide supported functionality only under a limited set of operating system or software packages (i.e., hardware drivers or accompanying software are not provided for all major operating systems). Although the relatively widespread MCU evaluation boards such as Arduinos can be used for the same purpose, many labs shy away from them because of uncertainty as to whether or not events can be marked with the required precision. The same holds for commercial USB input/output products for which no published timing test results are available (Wimmer et al., 2019).

With the present study, we aim to resolve this uncertainty and to demonstrate that USB-based "event trigger devices" that present themselves as serial (COM) ports can generally serve as adequate replacements for the previous gold standard parallel port. We describe the general principle underlying the commercial products that link the USB port to proprietary connectors expecting

a TTL pulse. Furthermore, we run a suite of latency tests on several such event trigger devices under two major operating systems (Windows, Linux) and compare the results against the performance of a traditional parallel port. We did not run tests under the macOS operating system because devices running macOS do typically not have native parallel port support. In the Supplemental Material, we provide a detailed tutorial on how to build event-trigger devices and software examples on how to operate them.

4.2 General principle underlying MCUs

A Microcontroller Unit (MCU) consists of at least one processor, memory (both volatile working memory and programmable memory for program code), and digital in- and output ports (e.g., for TTL signals or standardized buses like USB or the Serial Peripheral Interface (SPI)). Upon startup, the MCU reads the firmware from the internal memory and executes the contained instructions. The event-trigger devices analyzed here were programmed to read a trigger value from the PC, either directly via an embedded USB controller or a separate USB controller chip attached via an internal serial port, activate the corresponding output ports for a few milliseconds, and repeat these instructions from the beginning. The pseudocode for the simplest possible trigger device firmware consists of just a few lines (see Algorithm 1). This design involves at least two send- or receive-buffers (four in the case of a separate USB controller) which, combined with a fixed transfer rate and buffer sizes, put upper bounds on bandwidth (typically 14.4 kB/s at 115200 baud; 1.2 kB/s at 9600 baud) and lower bounds on latency (69.4 µs at 115200 baud with a single pair of buffers that can be flushed instantaneously up to 107 ms for two 64-byte buffers at 9600 baud). The simplest communication scheme translates one input byte to 8 bits, each of which is linked to one digital output and encoded in base 2 (e.g., the trigger 75 would translate to $0 \times 128 + 1 \times 64 + 0 \times 32 + 0 \times 16 + 1 \times 8 + 0 \times 4 + 1 \times 2 + 1 \times 1 = 01001011_2$, so the pins 7, 4, 2, and 1 would be enabled).

Algorithm 1 Example pseudocode for operating an event trigger device. Note that unlike in this example, some MCUs may have specialized functions that allow setting all output pins at once, without having to iterate through a loop.

```
while True do
    bits = read_byte_from_pc()
    output_pin = 0
    for bit_value in bits do
        set_output(output_pin, bit_value)
        output_pin = output_pin + 1
    end for
    wait_briefly()
    reset_outputs()
end while
```

Following the code in Algorithm 1, an event trigger (in form of a TTL pulse) can be caused by sending a single byte, the trigger value, to the device's input buffer. More functionality can be built in by replacing the MCU's firmware, implementing an instruction scheme that evaluates the input and sets outputs depending on the result of the computation. With minor adjustments, some MCUs can even be accessed via WebUSB to synchronize browser-based experiments with measurement devices via TTL pulses. Being able to program an MCU's firmware to execute almost arbitrary logic and communication protocols can thus be a big advantage. However, such firmware modifications are typically not possible for commercial devices, as these are often constrained by

built in, proprietary firmware. This constitutes the main difference between the MCUs tested here and other devices, such as the LabJack U3. Their general functionality and the communication protocol between the host PC and the trigger device stays identical.

4.3 Methods

To compare the latency and jitter of MCUs with that of the parallel port, we devised a test setup using a LabStreamer device (NeuroBehavioral Systems; Albany, CA, USA), a standard Dell Optiplex 750 desktop computer with a native parallel port (the host computer), a Teensy 3.2 MCU to simulate keypresses, and a LabJack U3 (LabJack Cooperation; Lakewood, CA, USA) and four other MCUs to be tested against the parallel port (see Figure 4.1). We established a communication protocol between the host computer and the LabStreamer device using the Lab Streaming Layer (LSL; https://github.com/sccn/labstreaminglayer). LSL is a software library for streaming timestamped measurement samples and TTL pulses over the local network. Clock offsets and uncertainty of the offset estimation between sender and receiver are periodically measured and subtracted out, so the receiver of a measurement sample (here: the LabStreamer) can record the time a TTL pulse was sent, rather than the time it was received.

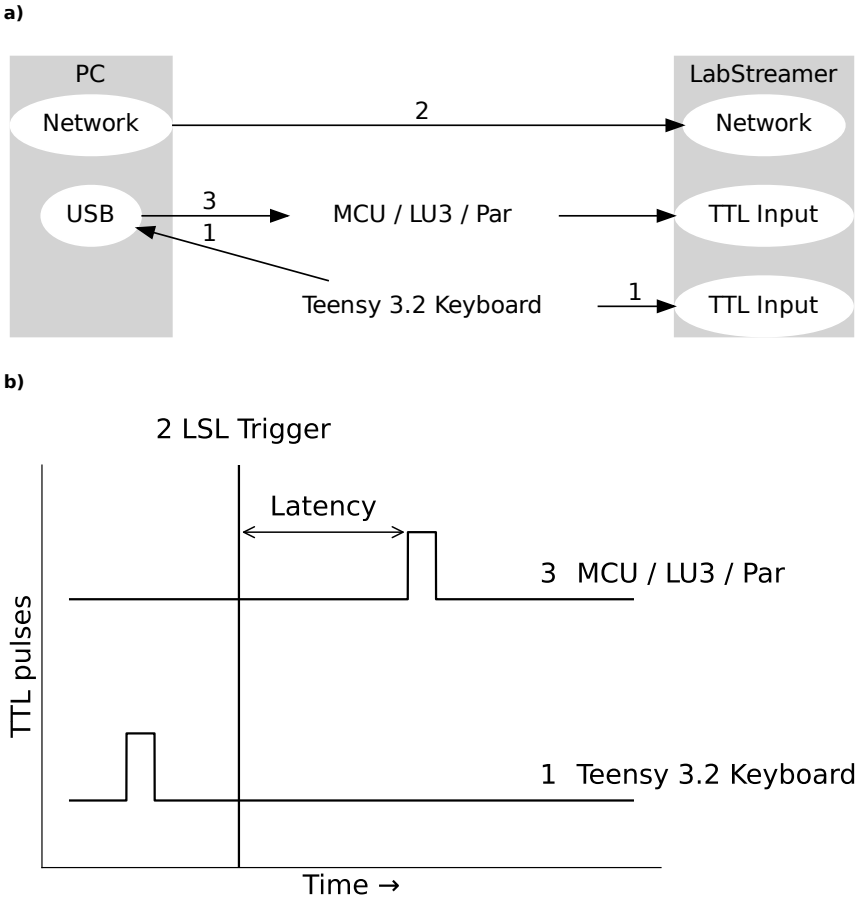

Figure 4.1. Testing setup. **a** Setup of the hardware. **b** Schematic representation of the signal time courses. For both panels, (1) the Teensy 3.2 keyboard simultaneously sends a key press to the PC and a TTL pulse to the LabStreamer; the PC receives the key press and sends (2) a timestamped LSL trigger via the network (marking timepoint zero), and (3) TTL pulse via the Microcontroller Unit (MCU), LabJack U3 (LU3), or parallel port (Par) to the LabStreamer. This yields the latency of the tested device (b).

The measurement process then proceeded as follows. The Teensy 3.2 MCU presented itself as a virtual keyboard to automatically send a keystroke to the host computer and simultaneously a TTL

pulse to the LabStreamer (sampling the digital inputs at 10 kHz) every 90 ms. Using LSL, the host computer and the LabStreamer then synchronized their clocks via a network connection. Upon receiving the simulated keypress signal, the host computer sent two additional signals: (i) a TTL pulse over the parallel port, LabJack U3, or a connected MCU, which was programmed to receive that signal via an emulated serial port running at 115200 baud per second, and (ii) an LSL trigger timestamped to the received keypress. The receiving MCU or LabJack U3 then read the signal byte and activated corresponding digital outputs, thus sending a TTL pulse to the LabStreamer (see Supplemental Material for the scripts). In each test run, at least 2500 measurements were made. Measurements with an LSL timestamp uncertainty below 0.01 ms, for example due to computational or network delays, were marked as invalid. Afterwards, we selected the first 2500 valid measurements from each test run for analysis.

We tested the LabJack U3 using two different methods of relaying the data: The standard `setFIOState` method, and the `writeRegister` method, which was recommended by the LabJack customer support to achieve faster latencies. Note that each method was tested on the same LabJack U3 device in separate measurement sessions.

Next to the parallel port and the LabJackU3, we tested four different, popular MCUs: (i) the widely used Arduino Uno as an example of an 8-bit MCU without native USB capabilities, (ii) the Arduino Leonardo (technically identical to the Arduino Pro Micro), an 8-bit MCU with an embedded native USB controller, (iii) the Teensy LC, an inexpensive 32-bit ARM MCU, and (iv) the Teensy 3.2, an affordable 32-bit MCU with enough communication interfaces and digital/analog in- and output ports for demanding experimental setups requiring the synchronization of several devices or processing of inputs, such as voice feedback. All devices (MCUs, parallel port, and LabJack U3) were tested with Python 3.6 and the Python bindings for the Psychtoolbox (Kleiner et al., 2007) on the Windows 7 and Ubuntu Linux 18.04 operating systems separately. To access the parallel port, the PsychoPy library was used (Peirce et al., 2019).

After data recording, we compared the device latencies of transmitting the byte from the PC to the LabStreamer by calculating the mean, median, standard deviation, and interquartile range of the data for each device and operating system. Additionally, we computed these statistics averaged across devices within each operating system group to see whether the devices performed better on one or the other operating system. All computations and plotting were done in Python using the libraries Numpy (van der Walt et al., 2011), Pandas (McKinney, 2010), Matplotlib (Hunter, 2007), and Seaborn (Waskom, 2021).

4.4 Results

Mean and median latencies as well as standard deviation and interquartile ranges are reported in Table 4.1 and visually presented in Figure 4.2. Across both operating systems, the parallel port had the lowest average latency and jitter. However, apart from the LabJack U3, all other tested devices consistently achieved latency and jitter comparable to that of the parallel port. For the LabJack U3, the mode of operation (`writeRegister` versus `setFIOState`) made a large difference. The `writeRegister` method resulted in smaller latencies, however this method also produced a large amount of outliers on Windows, but not on Linux (see Figure 4.2). The `setFIOState` method, in contrast, resulted in the longest latencies of all tested devices, but did not produce as many outliers as the `writeRegister` method. Notably, the LabJack U3 operated with the `setFIOState` method was much faster under Linux compared to Windows. Averaging over devices (excluding

Table 4.1. Latencies for all tested devices in milliseconds. The table includes an entry for the Teensy 3.2 device that simulated a keyboard in the test setup. Its latency is negative because the simulated keystroke happened prior to the timepoint zero (the LSL Trigger, cf. Figure 4.1).

Operating system	Device	Mean	SD	Median	IQR
Linux	Teensy 3.2 Keyboard	−1.650	0.410	−1.672	0.574
Linux	Parallel Port	0.212	0.032	0.211	0.05
Linux	Teensy 3.2	0.244	0.032	0.243	0.049
Linux	Teensy LC	0.254	0.032	0.255	0.052
Linux	LabJack U3 (`writeRegister`)	0.280	0.039	0.279	0.054
Linux	Arduino Leonardo	0.308	0.036	0.308	0.052
Linux	Arduino Uno	0.453	0.045	0.452	0.05
Linux	LabJack U3 (`setFIOStatus`)	0.616	0.120	0.603	0.057
Windows	Teensy 3.2 Keyboard	−1.756	0.334	−1.751	0.503
Windows	Parallel Port	0.210	0.030	0.209	0.049
Windows	Teensy 3.2	0.296	0.030	0.296	0.05
Windows	Teensy LC	0.319	0.033	0.319	0.05
Windows	Arduino Leonardo	0.352	0.032	0.353	0.05
Windows	LabJack U3 (`writeRegister`)	0.439	0.287	0.401	0.082
Windows	Arduino Uno	0.505	0.031	0.507	0.051
Windows	LabJack U3 (`setFIOStatus`)	1.144	0.099	1.157	0.158

Note. *SD* standard deviation, *IQR* interquartile range.

the LabJack U3 as an outlier) within each operating system group showed a small performance benefit of using Linux over Windows. Devices operated on Linux were 0.042 ms faster (mean). The device latency between operating systems was nearly identical for the parallel port (mean latency Linux = 0.212 ms, mean latency Windows = 0.210 ms). Looking at the remaining devices (again, excluding the LabJack U3 as an outlier), the smallest effect of operating system on device latency was observed for the Arduino Leonardo (mean latency Linux = 0.308 ms; mean latency Windows = 0.352 ms), and the largest effect for the Teensy LC (mean latency Linux = 0.254 ms; mean latency Windows = 0.319 ms).

4.5 Discussion

We evaluated four popular microcontrollers and a LabJack U3 device as potential replacements for the parallel port. All microcontrollers achieved latencies comparable to the parallel port in terms of both absolute latencies and jitter. The 8-bit MCUs (Arduino Uno, Arduino Leonardo) were generally slower than the 32-bit MCUs (Teensy LC, Teensy 3.2) but still consistently achieved latencies below 1 ms and below half a millisecond compared to the parallel port so even high-speed data recordings with sampling rates in the lower kilohertz range will therefore record triggers with a maximum delay and jitter of a few samples relative to a native parallel port. The LabJack U3 performed worse than all tested MCUs when accessed with code from the programming interface documentation (the `setFIOState` method). Direct access to the device's internal memory through the `writeRegister` method improved the latency, at the cost of increased complexity for the user. It is important to note, however, that when operated under Windows, the `writeRegister` method resulted in a number of large outliers of up to 7 ms. This problem did not occur when operating the LabJack U3 with the `writeRegister` method under Linux, or when operating it using the `setFIOState` method under either operating system. Apart from these eye-catching outliers in that particular case, the remaining devices did not produce many outliers. Although there were a few outliers larger than half a millisecond in two tests (Arduino Leonardo and Arduino Uno on Linux), they were well below the usual unavoidable timing differences, such as due to missed screen

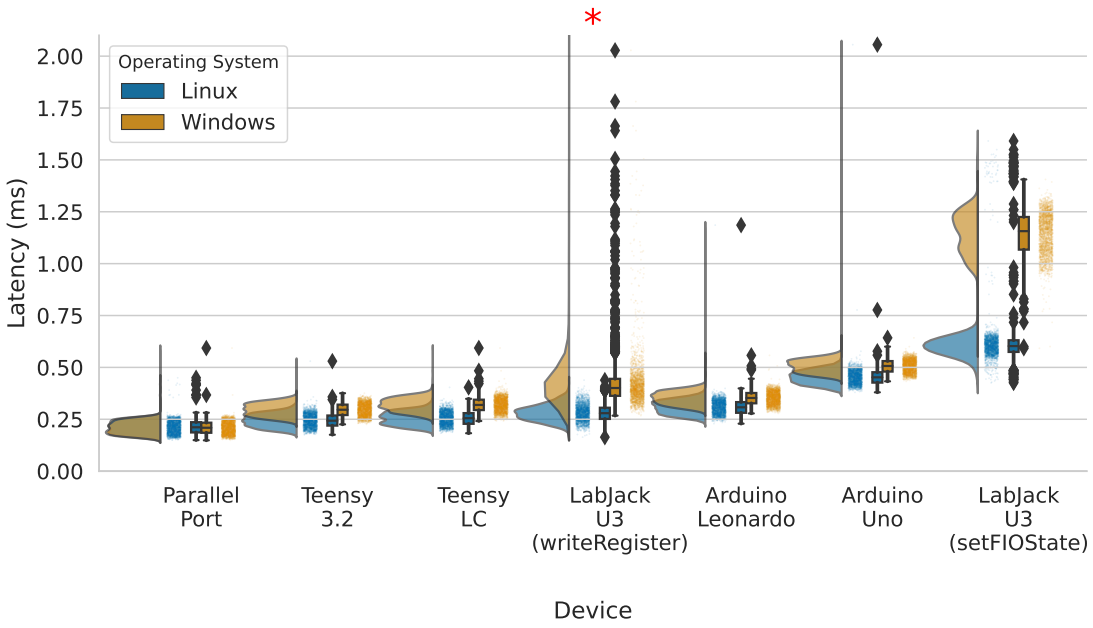

Figure 4.2. Raincloud plot (Allen et al., 2021) of TTL pulse latencies for the parallel port, all tested MCUs, and the LabJack U3. The LabJack U3 was tested using two different methods; (i) through the standard LabJack `setFIOState` method, and (ii) through writing data directly to the register with the `writeRegister` method. Under Windows, the LabJack U3 (`writeRegister`) had 11 more measurements between 2.3 and 7 ms that are not shown in the figure (mean = 4 ms), as indicated with the *red asterisk*.

refreshes and the latency of the input devices. Our testing setup also allowed us to measure the input lag due to the operating system without the underlying delays in real keyboards (see latency of Teensy 3.2 Keyboard in Table 4.1). This measured input lag provides a lower bound of unavoidable timing differences, that is, even if input (e.g., a keyboard or response box) and output (e.g., an event-trigger box based on an MCU) are hypothetically perfect and without delay, researchers still have to expect around 1.7 ms of delay with some jitter (see Table 4.1). This means that for example, in experiments marking a subject's response in a recorded electroencephalogram EEG by sending an event trigger after receiving a key press, the delay due to the operating system will be at least an order of magnitude above MCU TTL pulse latency (for a similar point, see Ulrich & Giray, 1989). Taken together, the results show that using MCUs (and with the mentioned caveats, the LabJack U3) to send digital event signals can reliably achieve adequate performance. Due to a lack of an Apple device with native parallel port support, we did not perform the tests under the macOS operating system. We speculate that the performance would have been slightly worse (longer latencies and increased jitter) compared to Windows and Linux, based on other results comparing the three operating systems (Bridges et al., 2020; Simpson, 2019). Yet for some macOS users, MCUs may still be an attractive option given the lack of alternatives.

Our findings build trust in the commercial products that have been introduced to the market in recent years to fill the gap left by parallel ports. However, the simplicity of our experimental setup and the devices we built for testing indicates that it may not be necessary to buy a commercial product: With some time and dedication, a replacement for the parallel port can be easily built (see Supplemental Material for instructions). Such Do-it-yourself (DIY) devices afford much greater flexibility than many commercially available devices: By choosing the MCU and other materials, as well as the software to run on the device, the end users themselves determine on which system the

device will run. For the MCU devices presented in this study, we supply firmware as well as high-level Python code under a permissive license in the Supplemental Material. The LabJack U3 on the other hand is shipped with a proprietary firmware and a dedicated software interface provided by the LabJack Cooperation. With the growing literature on open-source hardware in the spirit of open science (White et al., 2019), many other building instructions and software examples are available on the Internet. Unlike many commercial products, MCUs and the additional components needed to build a replacement for the parallel port are relatively cheap (around €30 in total) and are thus also available to researchers and users whose budgets do not allow them to buy expensive new hardware (Chagas, 2018). It is still important to stress that each laboratory setup must be appropriately tested and benchmarked before data recording (Plant et al., 2004; Plant & Quinlan, 2013; Plant & Turner, 2009). However, the same applies to commercial hardware and software. In summary, considering the affordable prices, ease of use, performance, and flexibility of MCUs with a native USB controller, there are many reasons for deploying them in laboratory studies.

4.6 Supplemental Material

The datasets generated and analyzed during the current study are archived and permanently available in the Zenodo repository under: https://doi.org/10.5281/zenodo.3838621.

The supplemental material including the analysis code and code used for data recording is available as a website: https://sappelhoff.github.io/usb-to-ttl. The source of the website is archived on Zenodo and permanently available under: https://doi.org/10.5281/zenodo.3838692.

The preprint for this chapter is available on PsyArXiv under: https://doi.org/10.31234/osf.io/x5dnb

4.7 Acknowledgments

This project was supported by the Fellow-Programm Freies Wissen of Wikimedia Germany, the Stifterverband, and the VolkswagenStiftung. We thank Susannah Goss for editorial assistance.

4.8 Author contributions

Following CRediT taxonomy: SA and TS: conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, validation, visualization, writing - original draft, writing - review & editing.

Corresponding author: Stefan Appelhoff

For a more detailed overview of author contributions, please see Appendix B.

4.9 Conflict of interest

TS has been provided with a pre-release version of the LabStreamer by NeuroBehavioral Systems for testing. The authors declare no other conflicts of interest.

4.10 Funding

Open Access funding enabled and organized by Projekt DEAL.

4.11 References

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., van Langen, J., & Kievit, R. A. (2021). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, *4*, 63. https://doi.org/10.12688/wellcomeopenres.15191.2

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online [Publisher: PeerJ Inc.]. *PeerJ*, *8*, e9414. https://doi.org/10.7717/peerj.9414

Canto, R., Bufalari, I., & DAusilio, A. (2011). A convenient and accurate parallel Input/Output USB device for E-Prime. *Behavior Research Methods*, *43*(1), 292–296. https://doi.org/10.3758/s13428-010-0022-3

Chagas, A. M. (2018). Haves and have nots must find a better way: The case for open scientific hardware [Publisher: Public Library of Science]. *PLOS Biology*, *16*(9), e3000014. https://doi.org/10.1371/journal.pbio.3000014

Clerc, M., Bougrain, L., & Lotte, F. (Eds.). (2016). *Brain-Computer Interfaces 2: Technology and Applications*. John Wiley & Sons, Inc. https://doi.org/10.1002/9781119332428

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Kleiner, M., Brainard, D., & Pelli, D. (2007). Whats new in Psychtoolbox-3?". *Perception 36 ECVP Abstract Supplement*.

Knight, D. (1997). Macs Need Parallel Ports. Retrieved April 22, 2020, from https://web.archive.org/web/20180626160731/https://lowendmac.com/1997/macs-need-parallel-ports/

Luck, S. J. (2014). *An introduction to the event-related potential technique* (Second edition). The MIT Press.

McKinney, W. (2010). Data Structures for Statistical Computing in Python, 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Plant, R. R., Hammond, N., & Turner, G. (2004). Self-validating presentation and response timing in cognitive paradigms: How and why? *Behavior Research Methods, Instruments, & Computers*, *36*(2), 291–303. https://doi.org/10.3758/BF03195575

Plant, R. R., & Quinlan, P. T. (2013). Could millisecond timing errors in commonly used equipment be a cause of replication failure in some neuroscience studies? *Cognitive, Affective, & Behavioral Neuroscience*, *13*(3), 598–614. https://doi.org/10.3758/s13415-013-0166-6

Plant, R. R., & Turner, G. (2009). Millisecond precision psychological research in a world of commodity computers: New hardware, new problems? *Behavior Research Methods*, *41*(3), 598–614. https://doi.org/10.3758/BRM.41.3.598

Simpson, S. (2019). Testing round trip USB Serial latency using Python. Retrieved February 19, 2021, from https://web.archive.org/web/20191227183540/http://blog.labhackers.com/?p=25

Stewart, N. (2006). A PC parallel port button box provides millisecond response time accuracy under Linux. *Behavior Research Methods*, *38*(1), 170–173. https://doi.org/10.3758/BF03192764

Ulrich, R., & Giray, M. (1989). Time resolution of clocks: Effects on reaction time measurement-Good news for bad clocks. *British Journal of Mathematical and Statistical Psychology*, *42*(1), 1–12. https://doi.org/10.1111/j.2044-8317.1989.tb01111.x

van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering*, *13*(2), 22–30. https://doi.org/10.1109/MCSE.2011.37

Voss, A., Leonhart, R., & Stahl, C. (2007). How to make your own response boxes: A step-by-step guide for the construction of reliable and inexpensive parallel-port response pads from computer mice. *Behavior Research Methods*, *39*(4), 797–801. https://doi.org/10.3758/BF03192971

Waskom, M. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

White, S. R., Amarante, L. M., Kravitz, A. V., & Laubach, M. (2019). The Future Is Open: Open-Source Tools for Behavioral Neuroscience Research [Publisher: Society for Neuroscience Section: Commentary]. *eNeuro*, *6*(4). https://doi.org/10.1523/ENEURO.0223-19.2019

Wimmer, R., Schmid, A., & Bockes, F. (2019). On the Latency of USB-Connected Input Devices. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3290605.3300650

5 | EEG-BIDS, an extension to the brain imaging data structure for electroencephalography

Cyril R. Pernet, Stefan Appelhoff*, Krzysztof J. Gorgolewski, Guillaume Flandin, Christophe Philips, Arnaud Delorme & Robert Oostenveld*

* *shared first authors*

Pernet*, C. R., **Appelhoff***, **S.**, Gorgolewski, K. J., Flandin, G., Phillips, C., Delorme, A., & Oostenveld, R. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Scientific Data*, *6*(1). https://doi.org/10.1038/s41597-019-0104-8

Abstract

The Brain Imaging Data Structure (BIDS) project is a rapidly evolving effort in the human brain imaging research community to create standards allowing researchers to readily organize and share study data within and between laboratories. Here we present an extension to BIDS for Electroencephalography (EEG) data, EEG-BIDS, along with tools and references to a series of public EEG datasets organized using this new standard.

5.1 Introduction

EEG was first applied in humans nearly a century ago (Berger, 1929). It records the electric potential fluctuations at the scalp, primarily from locally synchronous post-synaptic activity in the apical dendrites of pyramidal cells in the cortex. Widely used in both clinical and non-clinical settings, EEG is becoming increasingly important in cognitive neuroscience, with statistics from scientific reports showing that interest in EEG has been growing faster since the early 2000s. This can be attributed to interest in brain-computer interfaces and more sophisticated dynamics measures along with more accurate biophysical models for reconstructing sources. EEG is more versatile than other imaging modalities because (i) it is lightweight and requires relatively low-cost equipment, (ii) it can be used in many different environments (e.g., while sitting in a lab chair, driving, walking, playing a video game, sleeping, interacting with others in social situations, etc.), (iii) it can be used either alone or in conjunction with other imaging modalities, (iv) its task design constraints are less restrictive than metabolic (PET) or hemodynamic (fMRI) imaging methods, and (v) it captures neural activity with millisecond precision, making it possible to record cortical dynamics at the speed of perception, thought and action. Because of this versatility, the field of applications for EEG is broad. In turn, the commercial market for EEG systems is much larger than that of other imaging techniques (e.g., PET, MRI, MEG), resulting in a multitude of equipment manufacturers (more than 10 major manufacturers in neuroscience) building different hardware systems, usually with their own software and proprietary data formats. Manufacturers have little financial incentive to cooperate and provide compatible formats. This resulting diversity of formats is an impediment to reusing data as well as to building large-scale EEG databases.

The Brain Imaging Data Structure, originally proposed for Magnetic Resonance Imaging (MRI), is a human brain research community standard used for organizing and sharing brain imaging study data within and between laboratories for many (ultimately all) imaging modalities (Gorgolewski et al., 2016). BIDS primarily addresses the heterogeneity of data organization by following the FAIR principles (Wilkinson et al., 2016) of findability, accessibility, interoperability, and reusability. BIDS addresses findability and reusability by providing rich metadata in dedicated sidecar files and interoperability by using existing standard data formats. Accessibility is not directly addressed by BIDS, but by repositories that build on BIDS, such as OpenNeuro (https://openneuro.org). By stipulating how to structure data using naming conventions and dedicated metadata files to store dictionaries (`.json`) and data (`.tsv`), BIDS fosters interoperability and reuse of acquired datasets. Because BIDS data are structured, BIDS also addresses issues of reproducibility by allowing the creation of fully automated data analysis workflows.

Here we report on the extension of BIDS to EEG data. EEG-BIDS builds upon the MEG-BIDS extension (Niso et al., 2018) and is presented concurrently with the iEEG-BIDS extension covering human intracranial electrophysiology (Holdgraf et al., 2019). In this document we only highlight the main features of EEG-BIDS. The full documentation of the EEG-BIDS extension can be found via the link to the specification on the general BIDS website (https://bids.neuroimaging.io/).

5.2 EEG-BIDS Summary

The extension of BIDS to EEG data closely follows the general BIDS specification (see Figure 5.1). Each subject's data corresponds to a directory of raw data containing subdirectories for each session and data modality. This is accompanied by a `dataset_description.json` file containing generic information about the dataset and in the case of the EEG modality, a metadata file with

the suffix `eeg.json`. The `eeg.json` file exhaustively specifies among other metadata details of the experimental task and the EEG recording system. Optional directories include the `sourcedata` directory, which can be used to supply original non-formatted data. Furthermore, a `stimuli` directory and a `code` directory can be present to allow data conversion and preprocessing to be reproduced, as indicated in the original specification (Gorgolewski et al., 2016). Within each subject directory, the `eeg` subdirectory contains the EEG and metadata. For instance, for a single session study, `sub-XX` would have subdirectory `eeg` which contains EEG files using the naming pattern `sub-XX_task-YY_eeg.<extension>` corresponding to acquisitions of EEG data. In addition, `sub-XX_task-YY_channels.tsv` must be specified describing the parameters of the data acquisition and two extra files, `sub-XX_task-YY_electrodes.tsv` and `sub-XX_task-YY_coordsystem.json`, should be specified if the positions of the electrodes are known (see below).

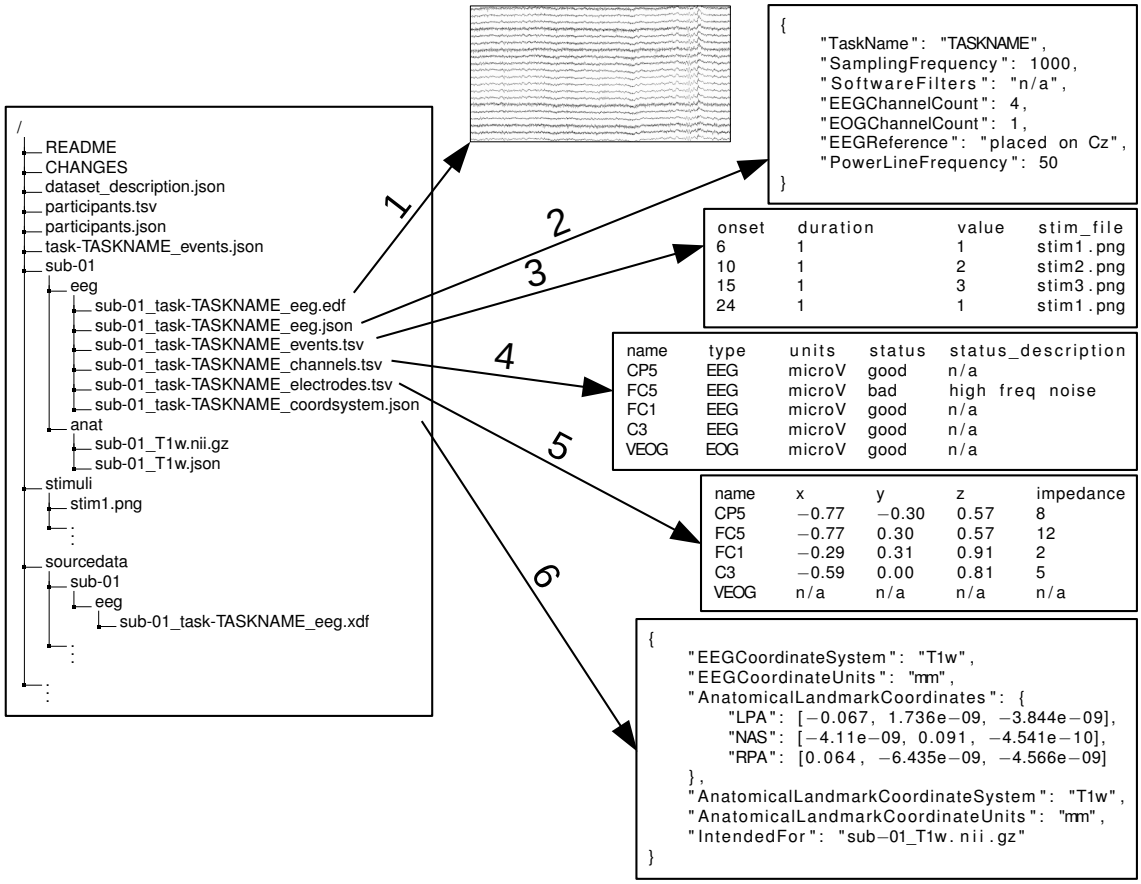

Figure 5.1. Exemplary EEG-BIDS dataset with previews of EEG files. The left side of the figure shows a standard BIDS directory tree with the root containing files describing the dataset in general (`README`, `dataset_description.json`), a file describing the participants (`participants.tsv`), and as several JSON files (`participants.json`, `task-TASKNAME_events.json`), which contain the description necessary to understand the contents of the TSV files. Note that JSON files at high levels get inherited by lower levels unless overridden (the "Inheritance Principle"). Next to the files at the root, there is a stimuli and a sourcedata directory that can be used to save the respective study data. Most important are the subject directories named `sub-<sub-label>` for each study participant. Nested in the subject directories are all recorded data split over modalities (`eeg` and `anat`, for the EEG and structural MRI data respectively). The right side of the figure shows the contents of the `eeg` modality directory, including the raw EEG data (1) and associated metadata (2). An `events.tsv` file (3) specifies all events that were recorded during the session and can reference presented stimuli with the `stimuli` directory of the dataset (see stim_file column). A `channels.tsv` file (4) provides further information about the raw EEG data and can contain information not present in the raw EEG data file such as filter settings and channel status (good/bad). Finally, an `electrodes.tsv` file (5) and an accompanying `coordsystem.json` file (6) provide electrode locations and specify which coordinate framework to use to interpret the electrode locations (for example with respect to a T1 weighted MRI scan).

As in the MRI specification of BIDS, `sub-XX_task-YY_events.tsv` files should be present, encoding all of the parameters of the experimental design (onset of events, trial type, duration, responses, etc.). While such information is often present as one or several binary "trigger" channels in the EEG recordings, the representation of events is rarely explicit in the original data (e.g., a numeric code is used to indicate the onset of a given picture presented in a given experimental condition) hence the necessity of these `events.tsv` files. Since the initial specification of BIDS for MRI in 2016, the Hierarchical Event Descriptor (HED) system for precise annotation of events has been integrated, which is particularly useful for electrophysiological data in which dynamics of brain activity are associated with multiple experimental events (Bigdely-Shamlo et al., 2016).

5.3 Specific EEG-BIDS Considerations

The process of converging on a list of suitable data formats for EEG-BIDS was governed by three major requirements: A suitable data format should (i) address the needs of a large portion of the global EEG community, (ii) be interoperable according to the FAIR principles (Wilkinson et al., 2016), and (iii) meet the technical requirements of neuroscientific workflows, such as saving numerical data with high precision.

As a solution to this challenge, the EEG-BIDS specification incorporates only two recommended "official" data formats: The European Data Format (EDF), which is an ongoing international effort to provide a common data format for electrophysiological recordings that began in 1992 (Kemp et al., 1992), and the BrainVision Core Data Format, developed by Brain Products GmbH. While the BrainVision Core Data Format was designed by Brain Products GmbH for its proprietary EEG recording equipment and analysis software, it is based on the Microsoft Windows INI file and has a concise documentation. Both of these formats follow the three requirements for suitable data formats for EEG-BIDS: (i) A recent survey indicates that they are widely used in the community (https://bids.berkeley.edu/news/bids-megeegieeg-data-format-survey), (ii) they have open access documentation and an open source implementation for both reading and writing in at least two programming languages that are widely used in the field (in this case, Python and MATLAB, among others), and (iii) they have high numerical precision (EDF: 16 bits, BrainVision Core Data Format: 32 bits). To accommodate a larger scientific audience and facilitate adoption, the EEG-BIDS standard also allows two "unofficial" commonly used data formats: The format used by the MATLAB toolbox EEGLAB (Delorme & Makeig, 2004, `.set` and `.fdt` files), and the Biosemi format (`.bdf`). While not actively encouraged, these two formats are included due to their popularity and their interoperability among the major software packages. Future versions of BIDS may extend the list of "officially" supported data formats, based on the fulfillment of the above mentioned three requirements for suitable data formats. Independently of the raw data format used, critical metadata about the recording are always available in BIDS `.tsv` and `.json` files.

In order to provide an unambiguous documentation of EEG data, we are clarifying two sets of terms that are often used interchangeably: Electrodes versus channels, and fiducials versus anatomical landmarks.

We distinguish between electrodes and channels using the following definitions: (i) An EEG electrode is a contact point attached to the skin, (ii) a channel is the combination of the analog differential amplifier and analog-to-digital converter that result in a potential (voltage) difference being stored in the EEG dataset. The "reference" and "ground" electrodes should in general not be

referred to as channels and only as electrodes. Some systems (e.g., Biosemi) have an active floating reference, whilst for most of the other systems, the potential at electrodes is neither amplified nor recorded. For EEG-BIDS, researchers must specify a `channels.tsv` file and may in addition specify an `electrodes.tsv` file with an accompanying `coordsystem.json` file.

Furthermore, we distinguish between fiducials and anatomical landmarks using the following definitions: (i) Fiducials are objects with a well defined location used to facilitate the localization of electrodes and co-registration with other geometric data such as the participant's own T1 weighted magnetic resonance head image, a T1 weighted template head image, or a spherical head model. Commonly used fiducials are vitamin E pills, which are clearly visible in an anatomical MR image, or reflective spheres that are localized with an infrared optical tracking system. (ii) Anatomical landmarks on the other hand define locations on a research subject such as the nasion, which is the intersection of the frontal bone and two nasal bones of the human skull. Fiducials are typically used in conjunction with anatomical landmarks.

5.4 Public EEG-BIDS Datasets

Several study examples (with zero-byte data files) are available in the BIDS-examples GitHub repository (https://github.com/bids-standard/bids-examples). We have also released three full datasets formatted using the EEG-BIDS standard:

- The Matching Pennies dataset (Appelhoff et al., 2018) is an example of a single recording session per participant. It was collected as part of a student project to replicate a brain-computer interface study of motor intention decoding.
- The Rishikesh dataset (Delorme, 2018) is an example of multiple recording sessions per participant. Participants were asked to meditate continuously whilst experience-sampling probe questions were presented at random intervals throughout the duration of the experiment.
- The simultaneous resting-state EEG-fMRI dataset (Clayden & Deligianni, 2016) offers resting-state data in EEG and fMRI modalities, and structural T1 weighted data and diffusion data (NODDI sequence).

5.5 Community Tools and Software Support

As part of the BIDS project, datasets formatted to follow the EEG-BIDS standard can be validated using the "bids-validator", a JavaScript application that runs locally as a command line version (using Node.js) or within an Internet browser (https://bids-standard.github.io/bids-validator/). With this validation tool, researchers can check their newly formatted datasets and make full use of the data structure's strengths for instance, checking for missing data or underspecified metadata.

The BIDS starter kit (https://github.com/bids-standard/bids-starter-kit) is a collection of community-driven guides, tutorials, helper scripts, and wiki resources to help researchers get started with BIDS. The resources cover two popular programming languages (Python and MATLAB) and will be extended over time to incorporate additional guides.

We are collaborating with the developers of the most widely used EEG data analysis tools in order to help EEG practitioners convert their existing data to the EDF or BrainVision Core Data Format. Data conversion utilities from many raw EEG data formats to the EDF and BrainVision format are available in MATLAB from the FieldTrip (Oostenveld et al., 2011) and EEGLAB (Delorme &

Makeig, 2004) toolboxes, and in Python from the MNE-Python (Gramfort et al., 2013) package. While specific software implementations are beyond the scope of the data specification, we here also want the reader to be aware of further work being undertaken by the open source community to interact with BIDS datasets. For instance, an EEGLAB "study" can now be exported as BIDS (`std_tobids.m`), FieldTrip can similarly export (`data2bids.m`) while SPM12 (Litvak et al., 2011) (`spm_bids.m`) and MNE-Python (in form of the MNE-BIDS project Appelhoff et al., 2019)) can read any BIDS dataset. Ultimately, reading/writing BIDS dataset will be fully automatic, as it is the case for MRI.

5.6 Data Analysis Pipelines and Beyond Sharing Raw Data

EEG's long history, versatility and variety of applications makes it a data- and method-rich technique. Recently, the OHBM Committee on Best Practice in Data Analysis and Sharing (COBIDAS) released a guideline for good practice and reproducibility in EEG (C. Pernet et al., 2020). According to the guideline, even the simplest processing pipeline already contains eight separate steps, making it clear that while EEG-BIDS helps with data sharing, it will remain non-trivial to develop automated processing pipelines of neurophysiology data such as already available for MRI data (Gorgolewski et al., 2017). On its own, EEG-BIDS is the first necessary step toward achieving validated and reproducible data analysis through standardizing the complete documentation of a dataset.

This article describes the new EEG extension for BIDS, and has limited itself to sharing raw data using previously developed community standards. Challenges that are specific to EEG, such as support for data formats, are still under active debate, and some additional formats will likely be incorporated once the technical issues are addressed and FAIR standards are achieved. The development of BIDS for EEG derivatives is also already underway (BIDS Extension Proposal 21): This will also make sharing of processed data possible, thereby fostering re-analyses, meta-analyses, and new analyses without the burden of data preparation.

5.7 Supplemental Material

The datasets generated and analyzed during the current project (i.e., the community survey) are available on GitHub under: https://github.com/sappelhoff/bids_format_survey.

The BIDS specification is available as a website: https://bids-specification.readthedocs.io. A PDF version of the specification is archived on Zenodo and permanently available under: https://doi.org/10.5281/zenodo.3686061.

The preprint for this chapter is available on PsyArXiv under: https://doi.org/10.31234/osf.io/63a4y

5.8 Acknowledgments

We thank the members of the BIDS community for their insights, stimulating discussions and helpful comments on early drafts of the EEG extension. We are grateful to Dora Hermes and Chris Holdgraf for sharing their article drafts of the iEEG specification and to Scott Makeig for discussions on data formats and metadata exchange. We would like to thank Deborah Ain for her very helpful comments and corrections.

5.9 Funding

This work was supported by the EU-H2020 Marie Curie "ChildBrain" Innovative Training Network grant no. 641652 to Robert Oostenveld.

5.10 Author contributions

C.R.P. Conception and design of the specification, moderating community interactions during the process, preparation of datasets and examples, coding of BIDS EEGLAB tools, making the figure, writing the manuscript.

S.A. Critical review of the specification, moderating community interactions during the process, preparation of datasets and examples, coding of the bids-validator extension, coding of BIDS MNE tools, making the figure, writing the manuscript, preparing the manuscript for resubmission.

K.G. Critical review of the specification, coding of the bids-validator, merging of BIDS specifications, critical review and final approval of the version submitted.

G.F. Critical review of the specification, coding of BIDS SPM and MATLAB tools, critical review and final approval of the version submitted.

C.P. Critical review of the specification, preparation of examples, critical review and final approval of the version submitted.

A.D. Critical review of the specification, preparation of datasets and examples, coding of BIDS EEGLAB tools, critical review and final approval of the version submitted.

R.O. Conception and design of the specification, moderating community interactions during the process, preparation of datasets and examples, coding of the bids-validator extension, coding of BIDS FieldTrip tools, critical review and final approval of the version submitted.

These authors contributed equally: Cyril R. Pernet and Stefan Appelhoff.

Corresponding authors: Cyril R. Pernet, Stefan Appelhoff, or Robert Oostenveld

For a more detailed overview of author contributions, please see Appendix B.

5.11 Conflict of interest

K.G. is a part-time Kaggle Research Consultant for Google LLC.

5.12 References

Appelhoff, S., Sanderson, M., Brooks, T. L., Vliet, M. v., Quentin, R., Holdgraf, C., Chaumon, M., Mikulan, E., Tavabi, K., Höchenberger, R., Welke, D., Brunner, C., Rockhill, A. P., Larson, E., Gramfort, A., & Jas, M. (2019). MNE-BIDS: Organizing electrophysiological data into the BIDS format and facilitating their analysis. *Journal of Open Source Software*, *4*(44), 1896. https://doi.org/10.21105/joss.01896

Appelhoff, S., Sauer, D., & Gill, S. S. (2018). Matching Pennies: A Brain Computer Interface Implementation Dataset. https://doi.org/10.17605/OSF.IO/CJ2DR

Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, *87*(1), 527–570. https://doi.org/10.1007/BF01797193

Bigdely-Shamlo, N., Cockfield, J., Makeig, S., Rognon, T., La Valle, C., Miyakoshi, M., & Robbins, K. A. (2016). Hierarchical Event Descriptors (HED): Semi-Structured Tagging for Real-World Events in Large-Scale EEG. *Frontiers in Neuroinformatics*, *10*. https://doi.org/10.3389/fninf.2016.00042

Clayden, J., & Deligianni, F. (2016). EEG, fMRI and NODDI dataset. https://doi.org/None

Delorme, A. (2018). BIDS formatted EEG meditation experiment data [type: dataset]. https://doi.org/10.5281/zenodo.2536267

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Gorgolewski, K. J., Alfaro-Almagro, F., Auer, T., Bellec, P., Capot, M., Chakravarty, M. M., Churchill, N. W., Cohen, A. L., Craddock, R. C., Devenyi, G. A., Eklund, A., Esteban, O., Flandin, G., Ghosh, S. S., Guntupalli, J. S., Jenkinson, M., Keshavan, A., Kiar, G., Liem, F., ... Poldrack, R. A. (2017). BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLOS Computational Biology*, *13*(3), e1005209. https://doi.org/10.1371/journal.pcbi.1005209

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., ... Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments [Number: 1 Publisher: Nature Publishing Group]. *Scientific Data*, *3*(1), 160044. https://doi.org/10.1038/sdata.2016.44

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python [Publisher: Frontiers]. *Frontiers in Neuroscience*, *7*(267). https://doi.org/10.3389/fnins.2013.00267

Holdgraf, C., Appelhoff, S., Bickel, S., Bouchard, K., DAmbrosio, S., David, O., Devinsky, O., Dichter, B., Flinker, A., Foster, B. L., Gorgolewski, K. J., Groen, I., Groppe, D., Gunduz, A., Hamilton, L., Honey, C. J., Jas, M., Knight, R., Lachaux, J.-P., ... Hermes, D. (2019). iEEG-BIDS, extending the Brain Imaging Data Structure specification to human intracranial electrophysiology. *Scientific Data*, *6*(1), 102. https://doi.org/10.1038/s41597-019-0105-7

Kemp, B., Värri, A., Rosa, A. C., Nielsen, K. D., & Gade, J. (1992). A simple format for exchange of digitized polygraphic recordings. *Electroencephalography and Clinical Neurophysiology*, *82*(5), 391–393. https://doi.org/10.1016/0013-4694(92)90009-7

Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R., Kilner, J., Barnes, G., Oostenveld, R., Daunizeau, J., Flandin, G., Penny, W., & Friston, K. (2011). EEG and MEG Data Analysis in SPM8. https://doi.org/10.1155/2011/852961

Niso, G., Gorgolewski, K. J., Bock, E., Brooks, T. L., Flandin, G., Gramfort, A., Henson, R. N., Jas, M., Litvak, V., T. Moreau, J., Oostenveld, R., Schoffelen, J.-M., Tadel, F., Wexler, J., & Baillet, S. (2018). MEG-BIDS, the brain imaging data structure extended to magnetoencephalography. *Scientific Data*, *5*, 180110. https://doi.org/10.1038/sdata.2018.110

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Intell. Neuroscience*, *2011*, 1:1–1:9. https://doi.org/10.1155/2011/156869

Pernet, C., Garrido, M. I., Gramfort, A., Maurits, N., Michel, C. M., Pang, E., Salmelin, R., Schoffelen, J. M., Valdes-Sosa, P. A., & Puce, A. (2020). Issues and recommendations

from the OHBM COBIDAS MEEG committee for reproducible EEG and MEG research. *Nature Neuroscience*, *23*(12), 1473–1483. https://doi.org/10.1038/s41593-020-00709-0

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

6 | General Discussion

In the introduction of this dissertation, we started with imagining somebody visiting a wine shop. That imaginary agent was faced with the tough challenge to buy a well tasting bottle of wine at a low price. If we were to directly apply the knowledge we gained from the chapters of this dissertation to that imaginary agent, we might say that (i) it is probably a good idea to allow the agent full control over which wines to sample, because it might boost their evidence processing (chapter 2), or (ii) that extremely bad and good tasting wines, or extremely cheap and expensive wines are likely overweighted compared to other wines, given an initial "default response" of the brain (chapter 3). However, such direct applications from basic research to constructed real-life examples are not always advisable[1], and furthermore it would be hard to apply what we learned about event marking using MCUs (chapter 4), or about neuroimaging data organization using BIDS (chapter 5) to a customer in a wine shop. Therefore, in the following sections I will not make any more references to the agent in the wine shop. I will summarize the key points from the past four chapters, and delineate open questions and future directions.

6.1 Control over sampling is beneficial for subsequent choice

In chapter 2 we reported a new aspect of decisions from experience, where participants first sample the possible outcomes of choice options before deciding for one of the options and receiving a random draw thereof as a reward. We found that the encoding of sample information in this setting was enhanced – both in behavior and in EEG patterns – when participants were allowed to fully self-control the sampling process (Appelhoff et al., 2022). Crucially, we obtained our results through comparison with baseline conditions containing fully matched stimulus sequences. This design choice allowed us to preempt several critiques. For example, we could exclude that choices under self-control might have been easier, due to optional stopping when differences between options were maximal (Hertwig & Pleskac, 2010), because these differences were identical in the "no control" baseline condition. As another example, we could exclude that CPP measures of evidence accumulation were due to any specific characteristics of the sampling sequence: The observed steeper "ramp-up" of CPP signals during full control was *not* due to a potentially stronger occurrence of rare or extreme outcomes near the stopping point (e.g., Donchin, 1981; DuncanJohnson & Donchin, 1977), because – again – the same outcomes occurred in the "no control" baseline condition.

The analysis of our data also produced some findings that are potentially relevant for future research. For example, we found a significant encoding of a samples "extremity", that is, its distance from the median sample value. While extreme samples in decisions from experience are often labeled extreme due to their rarity as well as their value (e.g., Pleskac & Hertwig, 2014; Wulff

[1] (and certainly not in this case)

et al., 2018), there is also a considerable amount of research looking at samples' extremity based solely on their placement at the "edges" of a range of outcome values (e.g., 1 and 9 when the range is 1-9; Ludvig et al., 2018; Ludvig et al., 2014; Ludvig & Spetch, 2011). Revisiting this branch of research with a neurocognitive (EEG) perspective holds the potential to uncover mechanisms as for *why* edge cases hold such sway over behavioral choice.

Our neurometric analysis furthermore showed a general overweighting of "outlying" extreme values, together with a strong bias towards larger numbers (see also chapter 3). This appeared to have a direct impact on stopping, because sampling sequences ended significantly more often with large extreme values than expected by chance. A promising next step towards understanding optional stopping in decisions from experience might be to scrutinize existing datasets (such as the large meta-analytic dataset from Wulff et al., 2018), with the analysis method proposed in chapter 2, to see whether stopping at extreme values is a more general trend. To follow up, EEG studies could be designed to see which neurocognitive measures might predict optional stopping.

6.2 Compression in behavioral choice is not predicted by EEG signals

When forming a choice based on a sequence of previously sampled values, decision-makers are often found to non-linearly weight each value before integration. In compressive weighting, inlying values have a higher impact than they should have compared to a linear weighting, whereas in anti-compressive weighting, this is true for outlying values (e.g., Fechner, 1860; Spitzer et al., 2017; Tsetsos et al., 2012; Tversky & Kahneman, 1992). In chapter 3 we harnessed recent findings on *when* compression versus anti-compression is elicited in humans (Clarmann von Clarenau et al., 2022). Showing participants the same stimuli sequences, we subtly changed task instructions from an "averaging" instruction to a "comparison" instruction, and succeeded to replicate and extend findings by Clarmann von Clarenau et al. (2022): The same participants compressed the sequences under averaging instructions, and anti-compressed them under comparison instructions. Crucially, we recorded EEG data throughout and applied a neurometric approach based on RSA to *predict* compression or anti-compression in subsequent choice, based on neural patterns measured at the time of individual sample presentation (see also chapter 2 and Appelhoff et al., 2022; Luyckx et al., 2019; Spitzer et al., 2017). Contrary to our expectation, the neurometric approach predicted anti-compression under both task instructions – even when the subsequent choice indicated a *compression* of sample values (according to behavioral modeling). The results indicate that extreme values are over-processed in the brain by "default". Future research will have the challenging task to find the processes that lead to an eventual downweighting of extreme values in the brain, as to reflect how these values become underweighted in choice behavior.

6.3 Microcontroller units are a viable alternative to the parallel port

In chapter 4 we showed that – in the context of sending "event markers" as in EEG research – most modern microcontroller units (MCUs) can achieve a latency of less than 1 ms, with a low amount of temporal jitter (Appelhoff & Stenner, 2021). All tested MCUs were comparable in performance to the previous (now disappearing) gold standard, the parallel port. Therefore, when performing time critical research such as ERP research (Luck, 2014), MCUs can be relied upon to

not introduce major measurement inaccuracies.

Together with our report on latencies of the different devices, we supplied supplemental materials on how to build MCU-based interfaces, as well as how to write the software to operate them. Prior to our study, much of this knowledge was incomplete, hidden in lab-internal protocols or deeply buried in user forums such as the PsychoPy (Peirce et al., 2019) forum (https://discourse.psychopy.org/)[2]. Judging from personal experience and communications, many colleagues facing the problem of needing an alternative to the disappearing parallel port either (i) used a significant amount of time to find such "hidden" solutions, (ii) spent a large amount of money for commercial solutions, or (iii) gave up on the problem and reverted to old computer equipment that still supported the parallel port. We hope that having an openly accessible report with background information, timing measurements, and software and build instructions will save future researchers this time, in the spirit of open science and open hardware (Fecher & Friesike, 2014; White et al., 2019). Furthermore, our report may make this technology accessible for researchers who may not be able to afford the often pricey commercial solutions (Chagas, 2018).

We performed our study at a peculiar point in time. Many researchers still have research equipment interfacing with the parallel port, but commercially available personal computers are typically not shipped with a parallel port anymore. With the present study, we provide a temporary solution, yet future technology will most likely rely on other technologies: For example on network protocols that use time stamps which are synchronized at analysis time (e.g., using "Cristian's algorithm" Cristian, 1989), as is done in the popular Lab Streaming Layer (LSL) software (https://github.com/sccn/labstreaminglayer). Using network protocols and time stamps instead of TTL markers transmitted via wires would allow for wireless recordings that become important when running mobile brain/body imaging studies (MoBI; Jungnickel et al., 2019), for example by using EEG in virtual reality contexts. In any case, until we see what the future holds, our study provides some solid grounds to continue doing EEG research without a parallel port.

6.4 A community-developed data standard helps to improve reproducibility in science

The Brain Imaging Data Structure (BIDS) is a data standard that provides schemas on (i) how to organize files and directories, (ii) how to structure metadata, and (iii) which information and documentation to supply alongside a dataset. One important aspect of BIDS is its extensibility through community processes, so-called BIDS Extension Proposals (BEPs). In chapter 5, we summarized the results of such a BEP aimed at extending the BIDS format to cover EEG as a data modality (Pernet et al., 2019).

A data standard like BIDS addresses fundamental aspects of good research data management. It provides guidelines on how to organize, annotate, track, and store research data, and thus provides the foundation for reproducibility of scientific results (Borghi & Van Gulick, 2021; Gorgolewski & Poldrack, 2016; Poldrack et al., 2019). However, BIDS is more than "just a data standard"; it also represents a growing community that builds and maintains that standard. As of writing this paragraph, more than 250 people have contributed in different ways, reaching from minor corrections to the specification text up to major overhauls and extensions of the standard (see the Contributors List[3]). Within 3 years of the acceptance of the EEG extension proposal presented

[2] In fact, Tristan Stenner and I initially met on that forum, trying to address the problem we then wrote the paper on.

[3] https://bids-specification.readthedocs.io/en/stable/99-appendices/01-contributors.html

in this dissertation, the report (Pernet et al., 2019) has been cited more than 130 times (Google Scholar, retrieved 2022-03-11). Major community projects such as the EEGManyLabs initiative (Pavlov et al., 2021) and the EEGManyPipelines initiative (https://www.eegmanypipelines.org/) have adopted BIDS for their workflows. In the domain of neuroimaging, support for BIDS is present in nearly all commonly used research software projects. Furthermore, there are more than 70 openly accessible EEG datasets in BIDS format shared on the OpenNeuro database alone (https://openneuro.org/; Markiewicz et al., 2021), noting that this is most likely a vast underestimation of all EEG datasets that have been shared in BIDS format since the publication of our report. These numbers serve as an indicator for both (i) the adoption of the standard by the wider neuroimaging community, and (ii) the need that our standard addresses.

To remain successful in the future, BIDS needs to address several challenges. With a growing standard that is continuously being extended, the need for active community members and maintainers to steer and streamline these extensions becomes pressing. Yet, finding volunteers who have enough time and knowledge to navigate a complex specification that encompasses modalities from arterial spin labeling MRI over physiological data to EEG derivatives is not easy. To lower the level of complexity both users and maintainers of BIDS are facing, the steering group and maintainers are currently discussing different models to go forward. One idea is to modularize the specification into smaller entities such as (i) raw data, (ii) derived data, (iii) model specifications (e.g., computational models, statistical models), and (iv) other data such as processes on provenance tracking. Addressing these challenges early will ensure the sustainability of BIDS as the major data standard in the domain of neuroimaging.

6.5 References

Appelhoff, S., Hertwig, R., & Spitzer, B. (2022). Control over sampling boosts numerical evidence processing in human decisions from experience. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhac062

Appelhoff, S., & Stenner, T. (2021). In COM we trust: Feasibility of USB-based event marking. *Behavior Research Methods*, *53*(6), 2450–2455. https://doi.org/10.3758/s13428-021-01571-z

Borghi, J. A., & Van Gulick, A. E. (2021). Promoting Open Science Through Research Data Management [Publisher: arXiv Version Number: 1]. https://doi.org/10.48550/ARXIV.2110.00888

Chagas, A. M. (2018). Haves and have nots must find a better way: The case for open scientific hardware [Publisher: Public Library of Science]. *PLOS Biology*, *16*(9), e3000014. https://doi.org/10.1371/journal.pbio.3000014

Clarmann von Clarenau, V., Pachur, T., & Spitzer, B. (2022). *Over- and Underweighting of Extreme Values in Decisions from Sequential Samples* (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/6yj4r

Cristian, F. (1989). Probabilistic clock synchronization. *Distributed Computing*, *3*(3), 146–158. https://doi.org/10.1007/BF01784024

Donchin, E. (1981). Surprise!? Surprise? *Psychophysiology*, *18*(5), 493–513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x

DuncanJohnson, C. C., & Donchin, E. (1977). On Quantifying Surprise: The Variation of Event-Related Potentials With Subjective Probability. *Psychophysiology*, *14*(5), 456–467. https://doi.org/10.1111/j.1469-8986.1977.tb01312.x

Fecher, B., & Friesike, S. (2014). Open Science: One Term, Five Schools of Thought. In S. Bartling & S. Friesike (Eds.), *Opening Science* (pp. 17–47). Springer International Publishing. https://doi.org/10.1007/978-3-319-00026-8_2

Fechner, G. (1860). *Elemente der Psychophysik*. Breitkopf und Härtel. https://books.google.de/books?id=6rINAAAAYAAJ

Gorgolewski, K. J., & Poldrack, R. A. (2016). A Practical Guide for Improving Transparency and Reproducibility in Neuroimaging Research [Publisher: Public Library of Science]. *PLOS Biology*, *14*(7), e1002506. https://doi.org/10.1371/journal.pbio.1002506

Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: Why small samples? *Cognition*, *115*(2), 225–237. https://doi.org/10.1016/j.cognition.2009.12.009

Jungnickel, E., Gehrke, L., Klug, M., & Gramann, K. (2019). MoBIMobile Brain/Body Imaging. *Neuroergonomics* (pp. 59–63). Elsevier. https://doi.org/10.1016/B978-0-12-811926-6.00010-5

Luck, S. J. (2014). *An introduction to the event-related potential technique* (Second edition). The MIT Press.

Ludvig, E. A., Madan, C. R., McMillan, N., Xu, Y., & Spetch, M. L. (2018). Living near the edge: How extreme outcomes and their neighbors drive risky choice. *Journal of Experimental Psychology: General*, *147*(12), 1905–1918. https://doi.org/10.1037/xge0000414

Ludvig, E. A., Madan, C. R., & Spetch, M. L. (2014). Extreme Outcomes Sway Risky Decisions from Experience: Risky Decisions and Extreme Outcomes. *Journal of Behavioral Decision Making*, *27*(2), 146–156. https://doi.org/10.1002/bdm.1792

Ludvig, E. A., & Spetch, M. L. (2011). Of Black Swans and Tossed Coins: Is the Description-Experience Gap in Risky Choice Limited to Rare Events? (A. Sirigu, Ed.). *PLoS ONE*, *6*(6), e20262. https://doi.org/10.1371/journal.pone.0020262

Luyckx, F., Nili, H., Spitzer, B., & Summerfield, C. (2019). Neural structure mapping in human probabilistic reward learning (D. Lee, J. I. Gold, D. Lee, & M. Chafee, Eds.) [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *8*, e42816. https://doi.org/10.7554/eLife.42816

Markiewicz, C. J., Gorgolewski, K. J., Feingold, F., Blair, R., Halchenko, Y. O., 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*, e71774. https://doi.org/10.7554/eLife.71774

Pavlov, Y. G., Adamian, N., Appelhoff, S., Arvaneh, M., Benwell, C. S., Beste, C., Bland, A. R., Bradford, D. E., Bublatzky, F., Busch, N. A., Clayson, P. E., Cruse, D., Czeszumski, A., Dreber, A., Dumas, G., Ehinger, B., Giorgio, G., He, X., Hinojosa, J. A., … Mushtaq, F. (2021). #EEGManyLabs: Investigating the Replicability of Influential EEG Experiments. *Cortex*, S0010945221001106. https://doi.org/10.1016/j.cortex.2021.03.013

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pernet, C. R., Appelhoff, S., Gorgolewski, K. J., Flandin, G., Phillips, C., Delorme, A., & Oostenveld, R. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography [Number: 1 Publisher: Nature Publishing Group]. *Scientific Data*, *6*(1), 103. https://doi.org/10.1038/s41597-019-0104-8

Pleskac, T. J., & Hertwig, R. (2014). Ecologically rational choice and the structure of the environment. *Journal of Experimental Psychology: General*, *143*(5), 2000–2019. https://doi.org/10.1037/xge0000013

Poldrack, R. A., Feingold, F., Frank, M. J., Gleeson, P., de Hollander, G., Huys, Q. J. M., Love, B. C., Markiewicz, C. J., Moran, R., Ritter, P., Rogers, T. T., Turner, B. M., Yarkoni, T., Zhan, M., & Cohen, J. D. (2019). The Importance of Standards for Sharing of Computational Models and Data. *Computational Brain & Behavior*, *2*(3), 229–232. https://doi.org/10.1007/s42113-019-00062-x

Spitzer, B., Waschke, L., & Summerfield, C. (2017). Selective overweighting of larger magnitudes during noisy numerical comparison. *Nature Human Behaviour*, *1*(8), 1–8. https://doi.org/10.1038/s41562-017-0145

Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal [ISBN: 9781119569107 Publisher: National Academy of Sciences Section: Biological Sciences]. *Proceedings of the National Academy of Sciences*, *109*(24), 9659–9664. https://doi.org/10.1073/pnas.1119569109

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/10.1007/BF00122574

White, S. R., Amarante, L. M., Kravitz, A. V., & Laubach, M. (2019). The Future Is Open: Open-Source Tools for Behavioral Neuroscience Research [Publisher: Society for Neuroscience Section: Commentary]. *eNeuro*, *6*(4). https://doi.org/10.1523/ENEURO.0223-19.2019

Wulff, D., Mergenthaler-Canseco, M., & Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*, *144*(2), 140–176. https://doi.org/10.1037/bul0000115

Appendices

A | List of manuscripts

* *shared first authors*

- **Appelhoff**, **S.**, Hertwig, R., & Spitzer, B. (2022). Control over sampling boosts numerical evidence processing in human decisions from experience. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhac062
- **Appelhoff**, **S.**, Hertwig, R. & Spitzer, B. (2022). EEG-representational geometries of psychometric distortions in approximate numerical judgment. (submitted to *PLOS Computational Biology*)
- **Appelhoff**, **S.**, & Stenner, T. (2021). In COM we trust: Feasibility of USB-based event marking. *Behavior Research Methods*, *53*(6), 2450–2455. https://doi.org/10.3758/s13428-021-01571-z
- Pernet*, C. R., **Appelhoff***, **S.**, Gorgolewski, K. J., Flandin, G., Phillips, C., Delorme, A., & Oostenveld, R. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Scientific Data*, *6*(1). https://doi.org/10.1038/s41597-019-0104-8

B | Declaration of contributions

Declaration pursuant to Sec. 7 (3), fourth sentence, of the Doctoral Study Regulations regarding my own share of the submitted scientific or scholarly work that has been published or is intended for publication within the scope of my publication-based work.

I. General information:

1. Last name, first name: Appelhoff, Stefan
2. Institutes: Fachbereich Erziehungswissenschaft und Psychologie & Max Planck Institute for Human Development
3. Doctoral study subjects: Psychology & Cognitive Neuroscience
4. Title: Master of Science (M.Sc.)

II. Numbered listings of works submitted (title, authors, where and when published and/or submitted):

1. **Appelhoff**, **S.**, Hertwig, R., & Spitzer, B. (2022). Control over sampling boosts numerical evidence processing in human decisions from experience. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhac062
2. **Appelhoff**, **S.**, Hertwig, R. & Spitzer, B. (2022). EEG-representational geometries of psychometric distortions in approximate numerical judgment. (submitted to *PLOS Computational Biology*)
3. **Appelhoff**, **S.**, & Stenner, T. (2021). In COM we trust: Feasibility of USB-based event marking. *Behavior Research Methods*, *53*(6), 2450–2455. https://doi.org/10.3758/s13428-021-01571-z
4. Pernet*, C. R., **Appelhoff***, **S.**, Gorgolewski, K. J., Flandin, G., Phillips, C., Delorme, A., & Oostenveld, R. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Scientific Data*, *6*(1). https://doi.org/10.1038/s41597-019-0104-8 (* *shared first authors*)

III. Explanation of own share of these works:

To characterize the types of contributions, I am following the CRediT taxonomy, which can be obtained from https://credit.niso.org/. The amount of contribution is declared on the following scale: "all - in vast majority - in majority - in large parts - in parts - equally with <co-author(s)>". The data is organized as: `<contribution type> (<amount>); ...`

re: II.1 Conceptualization (in large parts); Data curation (all); Formal analysis (all); Investigation (all); Methodology (in large parts); Project Administration (in large parts);

Software (all); Validation (all); Visualization(all); Writing - original draft (in majority); Writing - review & editing (in large parts)

re: II.2 Conceptualization (in parts); Data curation (all); Formal analysis (in majority); Investigation (all); Methodology (in parts); Project Administration (in large parts); Software (all); Validation (all); Visualization(in majority); Writing - original draft (equally with Bernhard Spitzer); Writing - review & editing (equally with Bernhard Spitzer)

re: II.3 Conceptualization (equally with Tristan Stenner); Data curation (in parts); Formal analysis (equally with Tristan Stenner); Investigation (equally with Tristan Stenner); Methodology (equally with Tristan Stenner); Project administration (in vast majority); Software (equally with Tristan Stenner); Validation (equally with Tristan Stenner); Visualization (in vast majority); Writing - original draft (in majority); Writing - review & editing (in majority)

re: II.4 Conceptualization (in large parts); Data curation (in large parts); Project administration (in large parts); Resources (in parts); Software (in vast majority); Supervision (in parts); Validation (in large parts); Visualization (all); Writing - original draft (in large parts); Writing - review & editing (in vast majority)

IV. Names, addresses, and e-mail addresses or fax numbers for the relevant co-authors:

Data is organized alphabetically (by first name) as: `<Name (e-mail)>; <affiliation>; <address>`

- Arnaud Delorme (adelorme@ucsd.edu); Swart Center for Computational Neuroscience, University of California San Diego, San Diego, California, USA; 9500 Gilman Dr, San Diego, CA 92093, USA
- Bernhard Spitzer (spitzer@mpib-berlin.mpg.de); Max Planck Institute for Human Development; Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany
- Christophe Phillips (c.phillips@uliege.be); GIGA Institute, University of Liège, Liège, Belgium; Cyclotron Research Centre, B30, 8, Allée du Six Août, 4000 Liège, Belgium
- Cyril R. Pernet (cyril.pernet@ed.ac.uk); Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, Scotland; Edinburgh Imaging Facility RIE, 57 Little France Crescent, Edinburgh, EH16 4TJ, The United Kingdom
- Guillaume Flandin (g.flandin@ucl.ac.uk); Wellcome Centre for Human Neuroimaging, London, United Kingdom; UCL Queen Square Institute of Neurology, Queen Square, London, WC1N 3BG, The United Kingdom
- Krzysztof J. Gorgolewski (krzysztof.gorgolewski@gmail.com); Google LLC; 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA
- Ralph Hertwig (hertwig@mpib-berlin.mpg.de); Max Planck Institute for Human Development; Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany
- Robert Oostenveld (r.oostenveld@donders.ru.nl); Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands; Radboud University

Medical Center (RUMC), P.O. Box 9101, 6500 HB, Nijmegen, The Netherlands

- Tristan Stenner (stenner@med-psych.uni-kiel.de); Institute of Medical Psychology and Medical Sociology, University Medical Center Schleswig Holstein, Kiel University, Kiel, Germany; Institut für medizinische Psychologie und Soziologie, PreuSSerstraSSe, 1-9, 24105 Kiel, Germany

V. Written confirmation of III. by co-authors (signatures):

- All co-authors have confirmed my contributions in written form (signatures).
- For reasons of data protection, the signatures of the co-authors are separately submitted to the Doctoral Examination Office (promotion@ewi-psy.fu-berlin.de), and are not directly included in this dissertation.

Stefan Appelhoff
Berlin, 12th of March, 2022

C | Declaration of independent work

I hereby declare that:

- I completed this doctoral thesis independently. Except where otherwise stated, I confirm that the work presented in this thesis is my own.
- Where information has been derived from other sources, I confirm that this has been indicated in the thesis.
- I have not applied for a doctoral degree elsewhere and do not have a corresponding doctoral degree.
- I have acknowledged the Doctoral Degree Regulations which underlie the procedure of the Department of Education and Psychology of Freie Universität Berlin, as amended on August 8th 2016.
- The principles of Freie Universität Berlin for ensuring good academic practice have been complied with.

Stefan Appelhoff
Berlin, 12th of March, 2022

D | Selection of other projects

Publications in academic journals

- MNE-BIDS: Organizing Electrophysiological Data into the BIDS Format and Facilitating their Analysis
  - **Appelhoff**, **S.**, Sanderson, M., Brooks, T. L., van Vliet, M., Quentin, R., Holdgraf, C., Chaumon, M., Mikulan, E., Tavabi, K., Höchenberger, R., Welke, D., Brunner, C., Rockhill, A. P., Larson, E., Gramfort, A., & Jas, M. (2019). Mne-bids: Organizing electrophysiological data into the bids format and facilitating their analysis. *Journal of Open Source Software*, *4*(44), 1896. https://doi.org/10.21105/joss.01896
- #EEGManyLabs: Investigating the Replicability of Influential EEG Experiments.
  - Pavlov, Y. G., Adamian, N., **Appelhoff**, **S.**, Arvaneh, M., Benwell, C. S., Beste, C., Bland, A. R., Bradford, D. E., Bublatzky, F., Busch, N. A., Clayson, P. E., Cruse, D., Czeszumski, A., Dreber, A., Dumas, G., Ehinger, B., Ganis, G., He, X., Hinojosa, J. A., ... Mushtaq, F. (2021). #Eegmanylabs: Investigating the replicability of influential eeg experiments. *Cortex*, *144*, 213–229. https://doi.org/10.1016/j.cortex.2021.03.013
- Capturing the Nature of Events and Event Context using Hierarchical Event Descriptors (HED)
  - Robbins, K., Truong, D., **Appelhoff**, **S.**, Delorme, A., & Makeig, S. (2021). Capturing the nature of events and event context using hierarchical event descriptors (hed). *NeuroImage*, *245*, 118766. https://doi.org/10.1016/j.neuroimage.2021.118766

Professional services

- Organizer of the 18th annual Summer Institute on Bounded Rationality, Berlin, Germany (2019)
- Ad-hoc reviewer for the Journal of Open Source Software (JOSS)

Contributions to science-related open-source projects

- **MNE-Python** (https://mne.tools)

  Analysis software for electrophysiology data

  - Active developer since 2018
  - Participant in the Google Summer of Code program 2019

  - Maintainer of the MNE-BIDS sub-package
  - Member of the Steering Committee since 2021
  - Mentor in the "New Contributors" coding-sprint 2021

- **BIDS** (https://bids.neuroimaging.io)

  Community-developed standard for neuroimaging data

  - Developer of standard extension for EEG
  - Maintainer for the standard since 2019
  - Active software developer for BIDS validator, examples, and website
  - Panelist in community discussions in Open Science Room at the OHBM conference 2019, 2020, and 2021
  - Delivering talks and workshops, for example: liveMEEG Conference 2020, PET-BIDS Workshop Copenhagen 2020, Open and Reproducible Neuroimaging Workshop Oldenburg 2020, Brain Products EEG BIDS Workshop 2019

- **eeg_positions** (https://stefanappelhoff.com/eeg_positions)

  Software for computing standard EEG electrode positions

  - Active developer and maintainer

- **pybv** (https://pybv.readthedocs.io)

  Software for writing EEG data to the BrainVision format

  - Active developer and maintainer

- **pyprep** (https://pyprep.readthedocs.io)

  Software for pre-processing EEG data

  - Active developer and maintainer

E | Curriculum vitae

2018 - 2022 **PhD Student**
Doctor Rerum Naturalium (Dr. rer. nat.)
Freie Universität Berlin, Germany
Max Planck Institute for Human Development, Berlin, Germany

2017 - 2018 **Einstein Fellow**
Einstein Foundation Berlin, Germany

2015 - 2017 **Master Student**
Master of Science (M.Sc.): "Social, Cognitive, and Affective Neuroscience"
Freie Universität Berlin, Germany

2012 - 2015 **Bachelor Student**
Bachelor of Arts (B.A.): "Integrated Social and Cognitive Psychology"
Jacobs University Bremen, Germany